

Mon Jun 26 09:10:26 2000

us-09-214-478-1.rge

Shin-din Chen

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 23, 2000, 06:06:58 ; Search time 878.41 Seconds

(without alignments)
-980.090 Million cell updates/sec

Title: US-09-214-478-1

Perfect score: 885

Sequence: 1 atgactcgtccggcgttc.....atgactctaccatgtag 885

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pr1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_sts.*
- 14: gb_sy.*
- 15: gb_un.*
- 16: gb_vi.*
- 17: em_fun.*
- 18: em_hum1.*
- 19: em_hum2.*
- 20: em_in.*
- 21: em_om.*
- 22: em_or.*
- 23: em_ov.*
- 24: em_pat.*
- 25: em_ph.*
- 26: em_pl.*
- 27: em_to.*
- 28: em_sts.*
- 29: em_sy.*
- 30: em_v1.*
- 31: em_v2.*
- 32: gb_htg1.*
- 33: gb_htg2.*
- 34: gb_in1.*
- 35: gb_in2.*
- 36: em_ba1.*
- 37: em_ba2.*
- 38: em_hum3.*
- 39: em_hum4.*
- 40: gb_pr4.*
- 41: gb_htg3.*
- 42: gb_htg4.*
- 43: gb_htg5.*
- 44: gb_htg6.*

- 45: gb_htg7.*
- 46: em_htg1.*
- 47: em_htg2.*
- 48: em_htg3.*
- 49: em_hum5.*
- 50: gb_pl3.*
- 51: gb_pr5.*
- 52: gb_htg8.*
- 53: gb_htg9.*
- 54: gb_htg10.*
- 55: gb_htg11.*
- 56: gb_htg12.*
- 57: gb_htg13.*
- 58: gb_htg14.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	885	100.0	3189	5	A52460	A52460 Sequence 4
C 2	885	100.0	8299	5	AR009152	AR009152 Sequence
C 3	885	100.0	35935	16	ADRCOMGEN	M73260 Mastadenovirus
C 4	875.4	98.9	35937	16	ADRCG	J01917 Adenovirus
C 5	328.8	37.2	3119	16	S82508	S82508 9E4: orf2..
C 6	319.2	36.1	35100	16	AF108105	AF108105 Human ade
C 7	275.4	31.1	34125	16	AT12CGA	Y73487 Adenovirus
C 8	273.8	30.9	3498	16	ADREB4	X51800 Adenovirus
C 9	222.2	25.1	34214	16	ADRGENOME	L19443 Human adeno
C 10	65	7.3	34094	16	AB026117	AB026117 porcine a
C 11	65	7.3	34094	16	PAD237815	AJ237815 Porcine a
C 12	62.2	7.0	3028	31	PAORR8G	L43363 Mastadenov
C 13	62.2	7.0	34094	16	AF083132	AF083132 Porcine a
C 14	49.8	5.6	31323	16	CAU77082	U77082 Canine aden
C 15	48.4	5.5	30288	16	CAU55001	U55001 Canine aden
C 16	48.4	5.5	30536	16	CAV1GENOM	Y07760 Canine aden
C 17	37.8	4.3	34446	16	AF030154	AF030154 Bovine ad
C 18	36.4	4.1	10499	2	AE000423	AE000423 Escherich
C 19	36.4	4.1	74377	43	AC016797	AC016797 Homo sapi
C 20	36.4	4.1	74377	43	AC016797	AC016797 Homo sapi
C 21	36.2	4.1	55123	42	AC016343	AC016343 Homo sapi
C 22	36	4.1	3181	34	AB023583	AB023583 Drosophill
C 23	36	4.1	3654	35	AF125986	AF125986 Drosophill
C 24	36	4.1	9282	35	AF083334	AF083334 Antheraea
C 25	36	4.1	13644	42	AC014231	AC014231 Drosophill
C 26	36	4.1	13644	42	AL136360	AL136360 Homo sapi
C 27	36	4.1	149886	12	MMHC438N12	AF049850 Mus muscu
C 28	36	4.1	153442	54	AC015838	AC015838 Homo sapi
C 29	35.4	4.0	123937	42	AC008146	AC008146 Trypanoso
C 30	35.2	4.0	155083	41	AC007847	AC007847 Homo sapi
C 31	35	4.0	180009	44	AC011777	AC011777 Homo sapi
C 32	35	4.0	234131	54	AC022389	AC022389 Homo sapi
C 33	34.8	3.9	2041	16	HS1732	M11316 HSV-1 (KOS)
C 34	34	3.8	468	3	BOVAMEL	J02695 Bovine amel
C 35	34	3.8	53793	40	AC007875	AC007875 Homo sapi
C 36	34	3.8	133952	32	AL135905	AL135905 Homo sapi
C 37	34	3.8	309036	41	AC006760	AC006760 Caenorhab
C 38	33.8	3.8	1654	16	MEAPHOSRA	M89920 Measles vir
C 39	33.8	3.8	9207	16	X16567	X16567 Measles vir
C 40	33.8	3.8	9208	16	MEAMIBE	X16569 Measles v
C 41	33.8	3.8	15894	16	AB012948	AB012948 Measles v
C 42	33.8	3.8	15894	16	AB012949	AB012949 Measles v
C 43	33.8	3.8	184762	57	AC016946	AC016946 Homo sapi
C 44	33.8	3.8	221563	52	AC010746	AC010746 Homo sapi
C 45	33.6	3.8	149093	45	AC019293	AC019293 Homo sapi

ALIGNMENTS

Qy	561	gtgtgtgtgcagttactgtctgatttaagttagatcagggtgcgtgctgtgcccg	720
Db	568	GTGTGTGTGCTGCAGTTACTGTGCTGATTTAAGTAGATCAGGGTGCCTGCTGTGCCGG	609
Qy	721	aggaaagcgcccttatgtctgcggcggtgcgaatcatcgtcaggagaccactgccatg	780
Db	508	AGGACAAGCGCGCTTATGCTGCGGCGTGCAGTATCGCTGAGGAGACCACTGCCATG	544
Qy	781	ttgtattctctgcagcagcagcgcgcgccagcagttatttcgcgcgtcgtcagcac	840
Db	548	TTGTATTCTTCAGGACCGAGCGCGCGGAGCAGTATTATTCGCGCCTGCTGCAGCAC	489
Qy	841	caaccgcctatctgatgcagcagattatgactctaccccccatgtag	885
Db	488	CACCGCCCTATCCTGATGCACGATTATGACTTACCCCATGTAG	444
RESULT	2		
LOCUS	AR009152		
DEFINITION	Sequence 2 from patent US 5756283.	DNA	PAT
ACCESSION	AR009152		
VERSION	AR009152.1	GI:3967957	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 8299)		
TITLE	Wilson, J.M., Fisher, K.J. and Gao, G.		
	Method for improved production of recombinant adeno-associated viruses for gene therapy		

Db	3822	TTGCCAGCTGGTTAGGATGGTGGATGGCCCATGTTTAAATCAGAGGTTATATGG	3881
Qy	481	taccggaggtgggtgaattacaacatgccaaagaggaatgtttatgccagcggttt	540
Db	3882	TACCGGAGGTGGTGAATTAACAACATGCCAAAGAGTAATGTTATGCCAGCGTGT	3941
Qy	541	atgaggggtgcccacttaactacgtcgttggatgatggccacgtgggtttctgtg	600
Db	3942	ATGAGGGGTGCCACTTAATCTACCTCGCTGCTGGTGTGATGATGCCACGTGGTCTGTG	4001
Qy	601	gtcccgccatgagcttggatagacagccttgccactgtgggattttgaacaattattgtg	660
Db	4002	GTCCCGCCATGAGCTTGGATACAGCGCTTGCACTGGGATTTTGAACAATATTGTG	4061
Qy	661	gtgtgtgtcagcttactgtcgtgatttaagtgagatcaagggtgcgctgtgcccgg	720
Db	4062	GTGCTGTCTGCAGTACTGTCTGCTGATTAAGTATGATCAGGGTGCCTGCTGCCCCG	4121
Qy	721	aggacaaggccttatctgctggggcgtgcgaatcatcgtggggagaccactgcatg	780
Db	4122	AGACAAGCGCCTTATGCTGGCGCGTGGCAATCATCGCTGAGGACCACTGCCATG	4181
Qy	781	tgtattctcagacagcagggcgggcgagcagatttattcgcgcctgctgcagcac	840
Db	4182	TGTATTCTCGAGGACGCGCGCGGCGGACGAGTTTATTCGCGCCTGCTGCAGCAC	4241
Qy	841	caccgcctactcgtgacacattatgactactacccccatgtag	885
Db	4242	CACCGCCCTATCTCGATGACGATTTATGACTACCCCATGTAG	4286
RESULT	3		
ADRCOMPEN/C	ADRCOMPEN	35935 bp	DNA
LOCUS	Mastadenovirus h5 gene, complete genome.		VRL
DEFINITION	Mastadenovirus h5 gene, complete genome.		08-APR-1996
ACCESSION	W73260	M29878	
VERSION	M73260.1	GI:209842	
KEYWORDS	protein 5 precursor; protein IIIa; protein V; protein pVII; protein pVIII; protein pVIIII		
SOURCE	Human adenovirus type 5		
ORGANISM	Human adenovirus type 5		
REFERENCE	1 (bases 1 to 35935)		
AUTHORS	Chroboczek, J., Bieher, F. and Jacrot, B.		
TITLE	The sequence of the genome of adenovirus type 5 and its comparison with the genome of adenovirus type 2		
JOURNAL	Virology	186 (1), 280-285 (1992)	
MEDLINE	92087470		
FEATURES	Location/Qualifiers		
source	1..35935		
	/organism="Human adenovirus type 5"		
	/db_xref="taxon:28285"		
CDS	<11565..11297		
	/note="52/55 K protein gene"		
	/codon_start=2		
	/protein_id="AAA96406.1"		
	/db_xref="GI:209843"		
	/translation="SFNNHRTIVAREEVAIGLHMLWDFVSALEQNPNSKPLMAQLFL		
	IVQHSRNEAFRDALNIVPEGEWLLDLNLIQSVQERSLSADKVAANISMLS		
	LGRFYARKIYHTVPIDKVKIEGFYMRNALKVLTLSDDLGYRNERIHKAVSVSR		
	RELSDELHMSLQALAGSGDREAESYFDAGADLWAPSRRALEAGAGPLAVAP		
	ARAGNVGVEYEDDEDEYEDGEY"		
	12318..14075		
	/codon_start=1		
	/product="protein IIIa"		
	/protein_id="AAA96407.1"		
	/db_xref="GI:289094"		
	/translation="MMODATDPVRAALQSPGSLNSTDDWRQVMDRIMSLTARNPDA		
	FRQOPQANRLSAILAEVVPARANPTEKVLIVNALAENRAIRDEAGLYDALLQRY		
	ARYNSGVQTLNRLVGVREAVARQAOQGNLGSVALNAPLSQPNVPRGQD		
	YTNFYSALELWTEPQSEVYQSGPDYFQTSROGLQTLNLSQAKNLQGLWGYRAP		
	GDRTVSSLLTPNSRLLLLLTAPTDSDGSVSREDYVLGHLLTLRYAIGQARVDEHTFO		
CDS			
	/codon_start=1		
	/product="protein pVII"		
	/protein_id="AAA96410.1"		
	/db_xref="GI:209847"		
	/translation="MRAHRRRRRRSHRRMRGGILPLLIPLIAAAIGAVPGIASVALQ		
	AQRH"		
	18003..18755		
	/codon_start=1		
	/product="protein pVI"		
	/protein_id="AAA96411.1"		
	/db_xref="GI:209848"		
	/translation="MEDINFASLAPRHGSRPFMGWMDIGTSMGSGAFSGSLWSGI		
	KNGSVKNGSKANNSSTGMLRDLKAEQNFQFQVQDGLASGISGVVDLANQVQNK		
	INSLDPRPVEEPVAVETVSPGEGKRPDRRETLVTQIDPPSYEALQKGLP		
	TTNPAPMAIGVIGQHTPVLDPADTQOKPVLPGTAVVTVTPRSASLRRASGSP		
	RSILPVASGNQSTLNSIVGLGVQSLKRRFC"		
	<25819..26484		
	/note="100 K protein"		
	/codon_start=1		
	/protein_id="AAA96412.1"		
	/db_xref="GI:209849"		
	/translation="ESCIIGTELOQSPDEKSAAPGLKTPGLMTSAYLRFVPEDY		
	HAHEIRFYEDQSRPNAELTACVITQGHILGLOALNKARQELLKRGVYLDQPSG		
	EELNPIDPPPOYQQOPRALASODGTQKRAAAATHGGILGSGRGFGGFGGSH		
	DGRLEPRRGSGFRGVRNTVTLGRIPLAGAPEIGNRFRQHGYNLRSSGAAGTARSP		
	TOP"		
	26485..26514		
	/partial		
	/note="of a 33 K protein"		
	/number=1		
	26713..27083		
	/note="of a 33 K protein"		
	/number=2		
	27174..27857		
	/codon_start=1		
	/product="protein pVIIII"		
	/protein_id="AAA96413.1"		
	/db_xref="GI:454806"		
	/translation="MSKEIPTMYMSYQPMGLAAGAAQDYSTRINTNSAGPHWISV		
	NGIRAHNRILLEQAATITTPNNLRNRPAAALVQYESPAPTVVLPDQAQEVQMT		
	NSGAQAGGFHVRSPGOGITHTLTIRGQTLNDSVSSSLGRLGQGTQIGAGRP		
	SFTPQALITQTSSESPSGIGITLQFIEEFVPSVYVFNPFSGPGHYDQFIPNEDA		
	VKDSADGYD"		
	8367 a 10073 c 9761 g 7734 t		
	BASE COUNT		
	ORIGIN		

Query Match 100.0%; Score 885; DB 16; Length 35935;
 Best Local Similarity 100.0%; Pred. No. 3e-246;
 Matches 885; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgactacgtccggcgcacattggcagatgacactacgacacacacgacatcggtgtgt 60
 Db 34077 ATGACTACGTCCGGCGTTCATTGGCATGACACTACGACCAACACGATCTCGGTGTCT 34018

QY 61 cgggcactccgtacagtaggagtcctacotccctttgagacagaaacccgcgtacc 120
 Db 34017 CGGCGCACTCCGTACAGTAGGAGTCCTACCTCCCTTTTGACAGAAACCCGCGCTACC 33958

QY 121 atactgagagatccgctgctgcccgaatgtaacactttgacacattgacacacagctgagt 180
 Db 33957 ATACTGAGAGATCCGCTGCTGCCGAATTAACACTTTGACAAATGCACACGTGAGT 33998

QY 181 tacgtgcgaggtttccctgcagtggtggtattacgctgattcagaagaatgggttttccc 240
 Db 33897 TAGCTCGAGGTTCCCTCGAGTGGGATTTACGCTGATTCAGGAATGGTGTTCCTCC 33838

QY 241 tgggatatgttttaacgcgggagagctgttaactcctgaggaagtgtatgcagctgtgc 300
 Db 33837 TGGGATATGTTTAAACGCGGAGAGCTGTGAATCCTCAGAAAGTGTATCAGCTGTGC 33778

QY 301 ctgtgttgccaaatgatgatcatcagcagcagatgatccatggttaacagatccctgg 360
 Db 33777 CTGTGTGTGCCAAATGATGATATCATGATGACGAGCATGATCCATGTTACGATCCTGG 33718

QY 361 gcttccactgtcattgttcacgtcccggttccctgcagtgatagccgcggtgaggtt 420
 Db 33717 GCTCTCCACTGTCAATGTTCAGTCCCGCTCCCTGCAGTGTATAGCGCGGCGAGTT 33658

QY 421 ttggccagctggtttgagtagtggtgagtgagcgcctatggttaatacagagtttatgg 480
 Db 33657 TTGGCCACTGTTAGGATGCTGTGATGGCGCCCATGTTTAAATCAGAGTTTATATGG 33598

QY 481 taccgggaggtggtgaattacaacatgccaaagagagtaattgttatgtccacggtgtt 540
 Db 33597 TACCGGAGGTGTTGTAATTAACAATGCCAAAGAGTATGTTATGTCACGCTGTTT 33538

QY 541 atgaggggtgcacacttaactcgtcgtctgtgtgtatgtagccacgtgggtttctgtg 600
 Db 33537 ATGAGGGGTGCCACTTAATCTACCTGCGCTTGTGTGTATGATGCCACGTGGTCTGTG 33478

QY 601 gtccccccatgagctttggtatagcagcgccttcactgtgggttttgaacattatgtg 660
 Db 33477 GTCCCCCATGAGCTTGGATACAGCGCTTCACCTGTGGGATTTGAACAAATATTGTG 33418

QY 661 gtgtgtgtgcagttactgtgtgtatttaagttagatcagaggtgctgtgtgcccgg 720
 Db 33417 GTGCTGTGCTGCAGTTACTGTGCTGATTTAAGTGAATCAGGCTGCTGTGTCGCCGG 33358

QY 721 aggaacaggcccttatgtctgctggcggtgagaaatcactcgtcagagacacactgcatg 780
 Db 33357 AGGCAAGGCGCTTATGTCTGGCGGCTGGCAATCATCTGCTGAGAGACCACTGCCATG 33298

QY 781 ttgtattctcagagcagagcggcggtgagcagcttatttcgcgcctctcgtcagcac 840
 Db 33297 TTGTATTCTCAGGAGCGGCGGCGGCGGACAGCTTATTTCGCGCGCTGCTGCAGCAC 33238

QY 841 caccgcccctatcctgatgcacgattatgactctaccctccatgtag 885
 Db 33237 CACCGCCCTATCTGATGCACGATATGACTCTACCCCTCCATGTAG 33193

RESULT 4
 ADRCG/c ADRCG 35937 bp DNA VRL 14-MAR-1996
 LOCUS Adenovirus type 2, complete genome.
 DEFINITION J01917 J01918 J01919 J01920 J01921 J01922 J01923 J01924 J01925
 ACCESSION J01926 J01927 J01928 J01929 J01930 J01931 J01932 J01933 J01934
 J01935 J01936 J01937 J01938 J01939 J01940 J01941 J01942 J01943
 J01944 J01945 J01946 J01947 J01948 J01949 J01950 J01951 J01952

VERSION
KEYWORDS

J01953 J01954 J01955 J01956 J01957 K00086 K00394 K00395 K02367
 M13004 V00007 V00008 V00009 V00010 V00011 V00012 V00013 V00014
 V00015 V00016 V00017 V00018 V00019 V00020 V00023 V00024
 J01917.1 GI:209811

SOURCE

DNA polymerase; DNA-binding protein; RNA polymerase III;
 alternative splicing; coat protein; complete genome; genome-linked
 protein; glycoprotein; overlapping genes; polymerase; terminal
 repeat; unidentified reading frame; virus-associated RNA.
 Human adenovirus type 2.

ORGANISM

Viruses; GSDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 1 (sites)
 One, K. and Weissman, S.M.

REFERENCE

The nucleotide sequence of a low molecular weight ribonucleic acid
 from cells infected with adenovirus 2
 J. Biol. Chem. 246 (22), 6991-7009 (1971)

JOURNAL

2 (sites)
 Jornvall, H., Ohlsson, H. and Philipson, L.

MEDLINE

An acetylated N-terminus of adenovirus type 2 hexon protein
 Biochem. Biophys. Res. Commun. 56 (2), 304-310 (1974)

REFERENCE

74147480
 3 (sites)
 Celma, M.L., Pan, J. and Weissman, S.M.

JOURNAL

Studies of low molecular weight RNA from cells infected with
 adenovirus 2. I. The sequences at the 3' end of VA-RNA I
 J. Biol. Chem. 252 (24), 9032-9042 (1977)

MEDLINE

78046048
 4 (sites)
 Celma, M.L., Pan, J. and Weissman, S.M.

REFERENCE

Studies of low molecular weight RNA from cells infected with
 adenovirus 2. II. Heterogeneity at the 5' end of VA-RNA I
 J. Biol. Chem. 252 (24), 9043-9046 (1977)

JOURNAL

5 (bases 10514 to 10680)
 Pan, J., Celma, M.L. and Weissman, S.M.

MEDLINE

Studies of low molecular weight RNA from cells infected with
 adenovirus 2. III. The sequence of the promoter for VA-RNA I
 J. Biol. Chem. 252 (24), 9047-9054 (1977)

REFERENCE

78046050
 6 (bases 30812 to 30900)
 Zain, B.S. and Roberts, R.J.

JOURNAL

Characterization and sequence analysis of a recombination site in
 the hybrid virus Ad2+ND
 J. Mol. Biol. 120 (1), 13-31 (1978)

MEDLINE

78153757
 7 (sites)
 Akusjarvi, G. and Pettersson, U.

REFERENCE

Nucleotide sequence at the junction between the coding region of
 the adenovirus 2 hexon messenger RNA and its leader sequence
 Proc. Natl. Acad. Sci. U.S.A. 75 (12), 5822-5826 (1978)

JOURNAL

8 (bases 5986 to 6236)
 Ziff, E.B. and Evans, R.M.

MEDLINE

Coincidence of the promoter and capped 5' terminus of RNA from the
 adenovirus 2 major late transcription unit
 Cell 15 (4), 1463-1475 (1978)

REFERENCE

79084199
 9 (bases 21607 to 21816)
 Akusjarvi, G. and Pettersson, U.

JOURNAL

Sequence analysis of adenovirus DNA. I. Nucleotide sequence at the
 carboxy-terminal end of the gene for adenovirus type 2 hexon
 Virology 91 (2), 477-480 (1978)

MEDLINE

79119384
 10 (bases 1 to 156; 35804 to 35937)
 Arrand, J.R. and Roberts, R.J.

REFERENCE

The nucleotide sequences at the termini of adenovirus-2 DNA
 J. Mol. Biol. 128 (4), 577-594 (1979)

JOURNAL

79175344
 11 (sites)
 Zain, S., Sambrook, J., Roberts, R.J., Keller, W., Fried, M. and
 Dunn, A.R.

MEDLINE

Nucleotide sequence analysis of the leader segments in a cloned

TITLE	Nucleotide sequence of the EcoRI D fragment of adenovirus 2 genome		
JOURNAL	Nucleic Acids Res. 8 (10), 2173-2192 (1980)		
MEDLINE	81053687		
REFERENCE	24 (sites)		
AUTHORS	Akusjarvi,G., Mathews,M.B., Andersson,P., Vennstrom,B. and Petersson,U.		
TITLE	Structure of genes for virus-associated RNAI and RNAII of adenovirus type 2		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 77 (5), 2424-2428 (1980)		
MEDLINE	80234635		
REFERENCE	25 (sites)		
AUTHORS	Perricaudet,M., Le Moulec,J.M. and Petersson,U.		
TITLE	Predicted structure of two adenovirus tumor antigens		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 77 (7), 3778-3782 (1980)		
MEDLINE	81054654		
REFERENCE	26 (sites)		
AUTHORS	Anderson,C.W. and Lewis,J.B.		
TITLE	Amino-terminal sequence of adenovirus type 2 proteins: hexon, fiber, component IX, and early protein 1B-15K		
JOURNAL	Virology 104 (1), 27-41 (1980)		
MEDLINE	80237620		
REFERENCE	27 (sites)		
AUTHORS	Van Etten,R.A., Walberg,M.W. and Clayton,D.A.		
TITLE	Precise localization and nucleotide sequence of the two mouse mitochondrial rRNA genes and three immediately adjacent novel rRNA genes		
JOURNAL	Cell 22 (1 PT 1), 157-170 (1980)		
MEDLINE	81042299		
REFERENCE	28 (sites)		
AUTHORS	Persson,H., Jornvall,H. and Zabielski,J.		
TITLE	Multiple mRNA species for the precursor to an adenovirus-encoded glycoprotein: identification and structure of the signal sequence		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 77 (11), 6349-6353 (1980)		
MEDLINE	81101049		
REFERENCE	29 (sites)		
AUTHORS	Butner,W. and Veres-Molnar,Z.		
TITLE	Localization of the 3'-terminal end of the EcoRI B fragment-specific early mRNA of adenovirus type 2		
JOURNAL	FEBS Lett. 122 (2), 317-321 (1980)		
MEDLINE	81138857		
REFERENCE	30 (sites)		
AUTHORS	Akusjarvi,G., Zabielski,J., Perricaudet,M. and Petersson,U.		
TITLE	The sequence of the 3', non-coding region of the hexon mRNA discloses a novel adenovirus gene		
JOURNAL	Nucleic Acids Res. 9 (1), 1-17 (1981)		
MEDLINE	81150446		
REFERENCE	31 (bases 459 to 608)		
AUTHORS	Osborne,T.F., Schell,R.E., Burch-Jaffe,E., Berget,S.J. and		
Query Match	98.9%;	Score 875.4;	DB 16; Length 35937;
Best Local Similarity	99.3%;	Pred. No. 1.9e-243;	
Matches	879;	Conservative	0; Mismatches 6; Indels 0; Gaps
QY	1	atgactacgtcgcggcttcatttggcgtgacactacagcaacacgattctcggtgtctt	60
Ddb	34077	ATGACTACGTCGGCGCTTCATTTGGCATGACACTACGACCAACACGATCTCGGTTGCT	34018
QY	61	cggcgactcgtacagttaggagtcgtctacctcttttggagcagaacccggctacc	120
Ddb	34017	CGGGGCACCTCCGTACAGTAGGGATCGCCTACCTCTTTTGACAGACACCGCGCTACC	33958
QY	121	atactgaggatactcgtctgctgcccgaatgtatcactttgacaattgacaacgtgagt	180
Ddb	33957	ATACTTGGAGGATCATCGCTGCTGCCGGAATGTACACTTTGACAATGCACAACGTGAGT	33898
QY	181	tacgtgcgaggtctctccctcagttggtgatttaacctgattcagaatgggtgtccccc	240
Ddb	33897	TACGTGCGAGGCTCTCCCTCGAGTGTGGGATTTACGCTGATTTCAGGAATGGGTGTTCCC	33838
QY	241	tgggtatgtttctaacgcgggaggagctgttaactcctgaggaagtgtatgcacgtgtgc	300
Ddb	33837	TGGATATGTTCTTGACCGGGAGGAGCTTGTAATCCTTGAGGAAGTATGACACGTGTGC	33778

```

QY 301 ctgtgtgtgtgccaacattgatcatgacgagcatgatccatgcttaccagtcctgg 360
DQ 33777 CTGTGTGTGTGCAACATGATATCATGACGAGCATGATCATCGTTTACGAGTCTGTG 33718
QY 361 gcttcacactgcatgttccacgtcccggtccctcagtgatagccggcgaggtt 420
DQ 33717 GCTCTCCACTGTCTATTTCCAGTCCCGGTTCCTCAGTCAGTACGCGCGGCGAGGT 33658
QY 421 ttggccagctggtttaggatggtggtggtggtggtggtggtggtggtggtggt 480
DQ 33657 TTGGCCAGCTGTTTAGGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 33598
QY 481 taccggaggtgctgaattacaacatgccaaagaggttaattgcttccagcgtgttt 540
DQ 33597 TACCGGAGGTGTTGAATTACAACTGATGCCAAAGAGGTAATGTTTATGCTCAGCGTGT 33538
QY 541 atgaggggtgcacacttaattacactgctgctggtggtggtggtggtggtggtggt 600
DQ 33537 ATGAGGGGTGCGCACTTAATCTACCTGCGCTGTGTTGTTGTTGTTGTTGTTGTTGTT 33478
QY 601 gtcccccgcacatgctgtggtggtggtggtggtggtggtggtggtggtggtggtggt 660
DQ 33477 GTCCCGGCATGAGCTTTGGATGATGAGCGCTTGCACCTGTGGGATTTGAACAATATGTG 33418
QY 661 gtgctgtgctgactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 720
DQ 33417 GTGCTGTGCTGAGTACTGCTGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 33358
QY 721 aggaacagccttctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 780
DQ 33357 AGGACAGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 33298
QY 781 ttgtattctgctgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 840
DQ 33297 TTGTATTCTTCAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 33238
QY 841 caccgctctctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 885
DQ 33237 CACCGCCTATCTGATGACGATTAATGATCTATACCCCATGTAG 33193

RESULT 5
S82508
LOCUS
DEFINITION 3119 bp DNA VRL
ACCESSION 9E4: orf2. . orf7 [adenovirus type 9, Genomic, 6 genes, 3119 nt].
VERSION S82508.1 GI:1699390
KEYWORDS
SOURCE Human adenovirus type 9.
ORGANISM Human adenovirus type 9.
REFERENCE 1 (bases 1 to 3119).
AUTHORS Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
TITLE
JOURNAL
MEDLINE
REMARK
GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 178541] from the original journal article.
This sequence comes from Fig. 1.
Region: 9E4.
Location/Qualifiers
1. .3119
/organism="Human adenovirus type 9"
/db_xref="taxon:10527"
471. .848
/gene="E4"
/notes="early region 4"
471. .848
/gene="E4"
/notes="orf1 transforming protein; This sequence comes from Fig. 1"
/codon_start=1

```

```

/protein_id="AAB37504.1"
/db_xref="GI:1699391"
/translation="MAESLYAFIDSPGIAPOVQGTNRXTFFCPESFHPPHGVLL
HLKVSIVPTGYQGRFALNDYHARDILTDQSDVIFAGRRQLITVLLFNHDFLYVKK
GHPVGTLLERLVIFPSPKIALV"
889. .1281
/gene="orf2"
889. .1281
/gene="orf2"
/notes="This sequence comes from Fig. 1"
/codon_start=1
/protein_id="AAB37505.1"
/db_xref="GI:1699392"
/translation="MLORRGVSYHIVPGVLYVLEDFSTOMIKELPRFTHILEG
ITGDKRAYSSMOFLGASFGALAFSLASPTLSPGSELSAVVAODLSDFLOTLRRE
LRAGRTLLNLVNLTLQVVEQDLELL"
1278. .1631
/gene="orf3"
1278. .1631
/gene="orf3"
/notes="This sequence comes from Fig. 1"
/codon_start=1
/protein_id="AAB37506.1"
/db_xref="GI:1699393"
/translation="MKVCLIMKVEGALWELFMCGVDLHQQFVEIIQGMKNENYLMV
QCNLMIDIDGGPAFNVILMDVRVEPLEATVEHLENRVGFDLAVCFHQHSGERL
HLRDLHFIVLRDLRLE"
1634. .1996
/gene="orf4"
1634. .1996
/gene="orf4"
/notes="This sequence comes from Fig. 1"
/codon_start=1
/protein_id="AAB37507.1"
/db_xref="GI:1699394"
/translation="MVLPLPPPLNDROGSINMGWYRVLADVMRGIRMDGFFISS
DAEELLNLRWMTFSWTERQQRKDRRGICCSRAHFCWQKYDKVKRIHYNAND
SIOLAPSSISOGPFTTI"
1926. .2804
/gene="orf6"
1926. .2804
/gene="orf6"
/notes="This sequence comes from Fig. 1; protein sequence
mismatches (125[Q->R],126[T->A],132[S->A],147[H->R])"
/codon_start=1
/protein_id="AAB37508.1"
/db_xref="GI:1699395"
/translation="MTEIQSSLRHPYRRARLPRSDDETRASLTQHPILLPCDHA
DYNVSVRGLPACAGFTLLQEPFVPMILTPPEIKLKRCMSVCLCPATLDLVRAQ
MVGYSERWILHCHCSGSLQCCGTGTLISVWFRVYICMFMNQHPVTRQVLRNMP
KEIMYGSVEMRGRHLIYCIWYDGHVGSIIIFNMFSGSALNYGLNNVIMCCYCE
NMAIRMCRCARTRRLMKAVGLIIVRETCDPDPICSSRTSPRRQLRALMERHRI
LFSEYEVSRSSHSTRL"
2798. .3046
/gene="orf7"
2798. .3046
/gene="orf7"
/notes="This sequence comes from Fig. 1"
/codon_start=1
/protein_id="AAB37509.1"
/db_xref="GI:1699396"
/translation="ITVLDCEARLDFSEDGFISITDPLARQETVWIIDPKSSSRIN
ENISFKATRAERTYTVKAGGRLTRAGVKINKDT"
BASE COUNT 759 a 701 c 848 g 811 t
ORIGIN

Query Match 37.2%; Score 328.8; DB 16; Length 3119;
Best Local Similarity 63.1%; Pred. No. 9.3e-85;
Matches 526; Conservative 0; Mismatches 302; Indels 6; Gaps 1;
QY 46 cgatctcggtgtctcgccactcgtacagtaggagatcgctacacctctttgagaca 105

```

Mon Jun 26 09:10:26 2000

Db	1941	CAATCAGCTCGTCCGCCATCATCATATCGAGGCCCGCTTTACACGATCTGATGAG	2000
Qy	106	gaaccccgctacatactgagagatcatcgcgtgctgcgcgaatgaacacttgaca	165
Db	2001	GAACACGAGGCTCTACTGATGAACAACTCCCTGTTGCCGATGATGATCATGACAT	2060
Qy	166	atgcacacagctgagctgagagctgctcctgcagtggtggattacgctgattcag	225
Db	2061	TATCAATATGATGCTCTGCTGATGATGATGATGATGATGATGATGATGATGATG	2120
Qy	226	gaatgggtgtgttccctggatattggttgcacacattgattgattgattgattgatt	285
Db	2121	GAGTTCACGCTCCCTGGGATATGATGATGATGATGATGATGATGATGATGATGATG	2180
Qy	286	tgatgcacagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	345
Db	2181	TGATGTCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2240
Qy	346	ggttacagctcctgggtgctcctcctcctcctcctcctcctcctcctcctcctcct	405
Db	2241	GGGTACGAGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2300
Qy	406	gcggggggcaggttttggcagctgctgctgctgctgctgctgctgctgctgctgct	465
Db	2301	CGGGAGCGACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2360
Qy	466	cagaggtttatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct	525
Db	2361	CAGGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2420
Qy	526	atgtccagctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	585
Db	2421	ATGGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2480
Qy	586	cagcgggttctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct	645
Db	2481	CAGTGGGTTCATCATCCCAACATGAGCTTTGGCTGAGCGCCCTGAATATGAGCTG	2540
Qy	646	ttgaacaattgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	705
Db	2541	CTGAATAACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2600
Qy	706	cctctgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg	765
Db	2601	CGGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2660
Qy	766	gagac-----cactgctcattgttattcctgctgctgctgctgctgctgctgctgct	819
Db	2661	GAGACTTGGGATCCCGATCCATCTGACAGCGCCGACCGCCCGCGCGAGAGACTG	2720
Qy	820	attcgccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct	873
Db	2721	CTGAGGCGCTGATGAGAGGAGCAGACACCCATCCCTGTTTCCGAGTATGATCT	2774
RESULT	6		
AF108105/c			
LOCUS	AF108105	35100 bp	DNA
DEFINITION	Human adenovirus type 17 complete genome.		
ACCESSION	AF108105		
VERSION	AF108105.1	GI:4416335	
KEYWORDS	Human adenovirus type 17.		
SOURCE	Human adenovirus type 17		
ORGANISM	Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.		
REFERENCE	1 (bases 1 to 35100)		
AUTHORS	Chillon, M., Bosch, A., Zabner, J., Law, L., Armentano, D., Welsh, M.J. and Davidson, B.L.		
TITLE	Group D adenoviruses infect primary central nervous system cells more efficiently than those from group C		
JOURNAL	J. Virol. 73 (3), 2537-2540 (1999)		
MEDLINE	99139039		

REFERENCE	2 (bases 1 to 35100)
AUTHORS	Armentano, D. PhD and Souza, D.W.
TITLE	Direct Submission
JOURNAL	Submitted (20-Nov-1998) Molecular Biology, Genzyme Corporation, 31 New York Ave., Framingham, MA 01701-9322, USA
FEATURES	Location/Qualifiers
source	1. 35100
	/organism="Human adenovirus type 17"
	/db_xref="taxon:46922"
	complement(491..8266)
	/product="DNA polymerase"
	/codon_start=1
	/protein_id="AAD20327.1"
	/db_xref="GI:4416343"
	/translation="MOCVDTATNSPVEIKYHLKHALRLYEYNLRTPPDLDRDT
	MDSSRALVAFALRRPRAITWLPRLGLSVLEKPEQGESHAGHENHPGPPCLKF
	LLKGRVLYDEVQVORCEYCGREFYKHQEGSVRRRDFYFHHIHSNWMQEOQF
	PIGSHPTERLRFVYDYVETWMSFGKOLVPMFKFSGDPELVALADLAARLRW
	DWRERDPLFYCVTPKMAVQQQFRFLRDELQTLMAELWASFMQANPHLQWALRQ
	GLQCPEDLYEELKLPKHPKPRFMELIYVGNINGFDEIVLAQVINLRASVPGPF
	RITRNPFRAGKILFNDVTFALPNLSKKRTDFELWEGHGGCDDSDKFIQFLKVMRDT
	FALHTSLRKAQAYALPVEKGCPCYKAVNHFYMLGSYRADDGFRFLREIKWDDRYA
	LNRELWKGAGCYDIIRETLDYCAMDLVLAFAVLQDSYAHFIRDSVRPLPAHFN
	IFQPTISSNSHALFRQIVFAEQPQRTNFGPLGLGSPHSELYEYVRASRGGRCYFN
	IGIFSEPTIYDYGCMYASALTHPMKGPPLNRYRAMAAREWQMAADSDADYDFD
	KKCPGFTTIDADPPDEHLDVLPFCSRGKGLCWTNEPLRGEVATSDVLTLNHRG
	WRVIVPDERITVPEWKCAREYVQUNIAAKERADBDKNOTMRSIAKLNLALYGSF
	ATKLDNRKIVFSQDMSLLKSAAGANIKSSFLTDNLNLSAEYPALELRYLPOOL
	ALVDNDAESEDHRPAPFPYTPSGTGHVAYTYKPTFLDABEGDMLTLYSEKVDPL
	VNDNRYPSHVASFVLAWTRAFVSWSEFLYEEDRGTPLDQRPKISVYGDFTDSLFTER
	GRLMETRGKKRIKKGKLVFQPEQELTWLVECTVCAHCAAGADAPAFVFLAPKL
	YALQSLCPACGSSKGLKRAKHAAEALNLMVNCYLADAQGEDARSTSRMSLK
	RTLASAPGAPHPFTVTETTLTRILRPNKDMTALAADAHRLVPSRSRPNRNEVCWI
	EMP"
	complement(8311..10161)
	/codon_start=1
	/product="terminal protein"
	/protein_id="AAD20326.1"
	/db_xref="GI:4416342"
	/translation="MDYFLPLRNINRVEEPRASTTAAGITWMSRYLYGYRLMLED
	LAPGAPATRWPLRYQPPHFLVGYLRTICNDYVDSRAFRSLRYSEVVQGLQTV
	NWMSMANCTYTINTGAYHRFVDMDDQDITLRVQQAIAERAVADLALVQLRQVGT
	RWEDSASADDIETRLMHDYKYNLSCQCGQAWMAERLRIQAGQPDLLVLTALRLAN
	AVFNYSINRNSNVHRAATCLSLPCDCDWDADFLEFSDVDLDTLSTPQLIRCI
	VSALSIPNGDPHREMTGGVETLRPRGRGAVTETMRRRGEMIERFVDELVSRNRR
	RRAPPPPPPEEIEEYVMEDEDEAPGDFEREVATIAELIRLEDELTVSARNAQ
	FFNFAYFEAMERLEAGDSIMPLRHWKFFVTEHATTLNLTANLAFORLRYAVFAR
	HYELNLAQVYMRADADGVYISRVNNESSLGAFSOLMGRISNDLAAVRAEGDLQ
	EEIEIQFMSEIAYQDNGDVQVIELROPVNDADIEIDSVLSFRKVTGVPVTVORRQIQ
	DYNRRVAHASALRAQHRDLPERHADVPLPLPAGPEPLPPGARPRHR"
	10622..11746
	/codon_start=1
	/product="L155K"
	/protein_id="AAD20320.1"
	/db_xref="GI:4416336"
	/translation="MHPVLQMRPTTPATTATAVAGAGASQPOQTEMDLEEGSLA
	RLGAPSPRHPVQLQKQVRYPAYQANLFRDRSGEPEEEDRCFRAGREGLDR
	QVRLDEDFEPNEQTGTSAPARAAANLVAEQTVKQERNFQSFNNHRTLLARE
	EVAGLHMLDLAELVQNDKSLPTAQLFVQHSRONEAREALLAIEPEGRMLL
	ELINLIQSIVQERSLSLAEKVAAINYSVLGKFKYAKIYKTPVPIIDKEVIDSFQ
	MRNALVLTSDDLGVIRNDRIHKAVASRRRLSDRELMLSLRRLVGAAGGEESY
	FDMGADLHWQSPSRALANAYGPDLDDEEEEDAPAG"
	11769..13463
	/codon_start=1
	/product="L11a"
	/protein_id="AAD20321.1"
	/db_xref="GI:4416337"
	/translation="MQQADPAIRAAQLSQPSGLASDDWEAAMORIMALTTRNPSFGR
	QQQANRLSILEAVVPSRTNPKERVLAINALNAENKAIKPADEAGLYLTLALSERVGR
	YNTSNVQNLRLDITVDREAVARQKNEGLGSLVALNAFLATQANVPRGDDITIN
	FISALRLMTVEPQSVQSGPDYFFQTQGLQTVNLQAFKNLQGLMGVQAPVGR
	STVSSLLTPNSRLILLIAPFTDGSVGNRNSYLGHLTLTYREAIGAQVDEQTFQIT
	SVSRKAUGNDTDSLRLATNLNLLNRQKIPQAQYALSAREERILRYVQSVGLFMQEG

CDS	
CDS	
CDS	

ATPSAALDMTARNMPSYAAANRPFINKLMDYLHRAASMSDYFTNAILNPHWLPPG
 FYTEYDMPDNDGLWDDVSAFSPLOKROEAVRTPASEGAVGRSPPLSGLSLHS
 LPSVNSGRVSRPRLGLDEVLNDSLLQPPRVKNANNGIESLVKLNKRWKTYAODH
 REPAPARRQRHQRGLVWDDDESADSSVLDLGGSGVNPISRILQKLGRMF"
 13517..15070
 /codon_start=1
 /product="penton base"
 /protein_id="AAD20322.1"
 /db_xref="GI:4416338"

CDS

/translation="MRAVYSSPPPSYBVSMAQATLEVPFPPRYMABTEGNSIRY
 SELAPLDTTRVYLVDNKSADIASLANYQNDHSNLTIVVONNDFTPAEASTQTINFE
 RSWGGDLKTLITNPNVNEFTSKFARVMARKHPQGEATDLSKDILEYEFWE
 FTLPENGSFTIEDLNNAILNYLQVGRQNGVLESDIGVFDSGRNFKLWDPTVKL
 VMPGVTEAFHDPVLPCCGVDFTESRLNLLGIRKKQFQEGFRIMYEDLEGNI
 PALLDPPKYLESKKLEALEANAANGPARGDSSVSREVEKAAEKELVIEPKODDT
 KRNYLIEGTMTLYRSWLSYTRDPENGVSOWILLTPDVTCCAEQVYWSLPMQ
 DPTVFRSTQVSNYPVGAELMPFRKSPFNLANVYSLIRSYTSLTVFNKFPDQI
 LCRPPATITTVSENPALTDHGTILPLRSIRGVQVIVIDARRTCYVYKALGIVA
 PVLSSRTT"
 17009..17713
 /codon_start=1
 /product="pvi"
 /protein_id="AAD20323.1"
 /db_xref="GI:4416339"

CDS

/translation="MEDINFASLAPRHGTRPMTWNEIGTSQNLNGAFNWSVWSGL
 KNFSGLRTYIGNKANSSTGOLLREKLDQNFQKVVVDGLASGVVDIDANQVORE
 INSRDLPPTVEMEDAILPPKGRPRDAETILOVDEPPSYEEAVKAGMPTFR
 IYIPLATGVAKPATLDDLPPTPPAPKAAVPPVPPVATAVRRVPARQAQWQSTLH
 SVIGLVKSLKRRCY"
 20620..21249
 /codon_start=1
 /product="protease"
 /protein_id="AAD20324.1"
 /db_xref="GI:4416340"

CDS

/translation="MSGSSERELASIVDLCCGTFWAPTRQFPGLAGDKLACATV
 NTAGCTGVHWLAFWPNRSTCYMDFPFGSDRLKQIYSEYEAAMLRSAVASSP
 DRCLSEGTQDVGSAAGLFCMFLHAFVHPDRPMDGNPTNMLNGLVNGMLQ
 SPQVPLTRRQNEELYRELAHSPYFRSHRAIEHAFAEDKMKQLRVSQ"
 30935..32035
 /codon_start=1
 /product="fiber"
 /protein_id="AAD20325.1"
 /db_xref="GI:4416341"

CDS

/translation="MSKRLRVDDFNVPYGVARNONIPLFPFVSDGFKNFPPG
 VLSKLADPTIANGDVSLSVGGGLTQESMTVDKAPLOLANNKLELYVYDPEV
 SANKLSIKVGHGLKLDKSGAGLKDILGLVLTGKIGTENLQNTDSSRGISLV
 RAREGLTFDNDGYLVANPKVYDTLTPTDTPSPNCKDKSKLTLVTRGSGILV
 ANVSLIVSGKYQYIDHATNPTLKSFKILFLDNKGVLLPSSNLDSTYWNFRSDNLT
 SEAYKNAVEPNPILVAPKPTTGSKKYARDIVYGNLYLGLAYQPVVIVKTFNEEADS
 AYSITFEVWKEATVAREFETTFSTYSIAQ"

BASE COUNT 7988 a 9877 c 9981 g 7250 t
 ORIGIN

Query Match 36.18; Score 319.2; DB 16; Length 35100;
 Best Local Similarity 62.48; Pred. No. 7e-82;
 Matches 520; Conservative 0; Mismatches 308; Indels 6; Gaps 1;
 Qy 46 cyatcgtggtgtcgcgcactccgtagcagtaggtagctgtactctctttgagaca 105
 Db 33171 CAATCCAGCTGGCTCCGCCATCATCATCGAGGCGCCGTTTACCATCTGATGAT 33112
 Qy 106 gaaacccgcgtaccatactaggagatcatccgtgctgcccgaatgaacatttgaca 165
 Db 33111 GAAACCAAGGCTTCACGACTGATCAACATCCCTGTTGCCGATTTGTGATCAGCAGAT 33052
 Qy 166 atgcacaacgtgagttacgtgcgaggtctccctgcagtggtgagttacgtgattcag 225
 Db 33051 TATCATATGATGTTCTGTCGCGGATACCATGATGCTGCTGGCTTTACCTGCTCCAA 32992
 Qy 226 gaatgggtttctcgtgggtagtgggttcttaacgcgggagagagctgttaacctctgaggaag 285
 Db 32991 GAGTTTCCAGTCCCTGGGATATGATCCTCCACACAGAGAAATAAATTTTAAAGA 32932

Qy 286 tgtatgcacgtgtgctgtgtgtgtcccaacattatcatcatcagagagatgatcatcat 345
 Db 32931 TGTATGTCAGTGTGCTGTGCCCGCTACCTGGACTTGGTAAGAGTTCAGATGGTGA 32872
 Qy 346 ggttacgagtcctggctgctccatctgtcattgtccacgtcccggttccctcagtgata 405
 Db 32871 GGGTACGAGCGGTGATCTGCAATGCCACTGTTTCATCCCTGCTCCCTGCAGTGTGCG 32812
 Qy 406 gcccgcggcgaggtttgtgcccagctggttttagtagtggtggtggtggtggtttaa 465
 Db 32811 CGGGGAGCACCTGCTGCGCGTGTGTTGTCAGGAGATCAATTTACGGGTGATGTTCAAC 32752
 Qy 466 cagaggtttatatgtgaccgggaggtggtggaattacaacatgccaagagaggttaatttt 525
 Db 32751 CAGCGCTTCCCTGCTGTCGCGCAGATGTTAAACAGAAACATGCCCCAAGAGATCATGTAT 32692
 Qy 526 atgtccacgctgtttatgaggggtgcgcaattaatctacctgctgtgtgtatgagtc 585
 Db 32691 ATGGCAGTGTGTTTCATGAGGGGCGAGCACCTGATATCTGCGGCATCTGTCAGATGGT 32632
 Qy 586 cactggt 645
 Db 32631 CACGTGGTGTCCATCATCCTCCCAACATGAGCTTTGGCTGGAGCACCTGATTAATGAGGCTG 32572
 Qy 646 ttgacaattgt 705
 Db 32571 CTGATTAACATGTTGATTTATGCTGCTGCTACTTACTGTGAGAACATGTCAGATCAGATG 32512
 Qy 706 cgt 765
 Db 32511 CGGTGCTGTGCTCGAGCAGCAGGACGATGATGCTGAGAGCTGTGCGGATCATAGTATGA 32452
 Qy 766 gacac-----cactgcaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 819
 Db 32451 GAGACTGTGATCCCGATCCCATTTTTCAGACGCGCAGCCGCGGCGGCGGAGAGCTG 32392
 Qy 820 attcgcgcgt 873
 Db 32391 CTGAGGCGCTGATGAGAGGACAGACCCATCTGTTTTCGAGTATGAATCT 32338

RESULT 7
 AT12CGA/c 34125 bp DNA VRL 17-FEB-1997
 LOCUS Adenovirus type 12 DNA, complete genome.
 DEFINITION X73487
 ACCESSION X73487
 VERSION X73487.1 GI:313361
 KEYWORDS complete genome; core protein; DNA polymerase; DNA-binding protein; endoprotease; fiber protein; hexon protein; large T-antigen; maturation protein; minor core protein; penton protein; periplental hexon-associated protein; promoter; repeat region; small t-antigen; transcriptional activation.
 SOURCE ORGANISM
 HUMAN adenovirus type 12.
 Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 REFERENCE 1 (bases 1 to 34125)
 Sprengel, J., Schmitz, B., Heuss-Neitzel, D., Zock, C. and Doerfler, W. Nucleotide sequence of human adenovirus type 12 DNA: comparative functional analysis
 J. Virol. 68 (3), 379-389 (1994)
 JOURNAL 94076430
 MEDLINE 2 (bases 1 to 34125)
 REFERENCE Sprengel, J.
 Direct Submission
 Submitted (21-JUN-1993) J. Sprengel, Institute of Genetics/Dept. Virology, Weyerstr. 121, 50931 Cologne 41, FRG
 JOURNAL 3 (bases 1 to 34125)
 REFERENCE Zock, C., Iselt, A. and Doerfler, W.
 A unique mitigator sequence determines the species specificity of the major late promoter in adenovirus type 12 DNA
 J. Virol. 67 (2), 682-693 (1993)
 JOURNAL 93124560
 MEDLINE

REMARK	(sites)	source
REFERENCE	4. (bases 4831 to 10470)	1. 34125
AUTHORS	Shu, L.M., Hong, J.S., Wei, Y.F. and Engler, J.A.	/db_xref="taxon:28282"
TITLE	Nucleotide sequence of the genes encoded in early region 2b of human adenovirus type 12	1. 161
JOURNAL	Gene 46 (2-3), 187-195 (1986)	/function="inverted terminal repeat"
MEDLINE	87106854	6
REFERENCE	5 (bases 1 to 3957)	/citation=[5]
AUTHORS	van Ormondt, H. and Galibert, F.	/citation=[6]
TITLE	Nucleotide sequences of adenovirus DNAs	/citation=[7]
JOURNAL	Cur. Top. Microbiol. Immunol. 110, 73-142 (1984)	/citation=[8]
MEDLINE	85002829	/citation=[12]
REFERENCE	6 (bases 1 to 3957)	/replace="a"
AUTHORS	Shinagawa, M. and Padmanabhan, R.	8
TITLE	Comparative sequence analysis of the inverted terminal repetitions from different adenoviruses	/citation=[5]
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 77 (7), 3831-3835 (1980)	/citation=[6]
MEDLINE	81054665	/citation=[7]
REFERENCE	7 (bases 1 to 3957)	/citation=[8]
AUTHORS	Sugisaki, H., Sugimoto, K., Takanami, M., Shiroki, K., Saito, I., Shimoto, H., Sawada, Y., Uemizu, Y., Desugi, S. and Fujinaga, K.	/citation=[12]
TITLE	Structure and gene organization in the transformed Hind III-G fragment of Ad12	/replace="t"
JOURNAL	Cell 20 (3), 777-786 (1980)	9. 407
MEDLINE	81022638	/note="crossref Ad12 EPD30061 and EPD11197"
REFERENCE	8 (bases 1 to 3957)	/function="Ela promoter region (-499 to +100)"
AUTHORS	Tolun, A., Alestrom, P. and Pettersson, U.	369
TITLE	Sequence of inverted terminal repetitions from different adenoviruses: demonstration of conserved sequences and homology between SA7 termini and SV40 DNA	/citation=[5]
JOURNAL	Cell 17 (3), 705-713 (1979)	/citation=[6]
MEDLINE	80001962	/citation=[7]
REFERENCE	9 (bases 1488 to 3861)	/citation=[8]
AUTHORS	Kimura, T.	/citation=[12]
TITLE	Structure and sequence analysis of the transforming region ElB of human adenovirus type 12	/replace=" "
JOURNAL	Sapporo Iseku Zasshi 52, 253-267 (1983)	503. 1099
REFERENCE	10 (bases 20966 to 22966)	/note="EIA; transcription activation: early protein; alternative splicing; crossref SWISS-PROT:Ela_ADE12, P03259"
AUTHORS	Kruijer, W., van Schaik, F.M., Speijer, J.G. and Sussenbach, J.S.	/codon_start=1
TITLE	Structure and function of adenovirus DNA binding protein: comparison of the amino acid sequences of the Ad5 and Ad12 proteins derived from the nucleotide sequence of the corresponding genes	/protein_id="CAA51877.1"
JOURNAL	Virology 128 (1), 140-153 (1983)	/db_xref="GI:313362"
MEDLINE	83277521	/db_xref="SWISS-PROT:P03259"
REFERENCE	11 (bases 24334 to 24703)	/translation="MRTMTPLVLSYQEAADLLEHLVDNFNEVPDSDLLYVPSLYEL YDLDVSGEDNNEQAVNEFFESLILASEGLFPEPVLSPVCEPIGECMPOLHP EDMDLICYEMGFCSDSEDEQENAHVSAASAAADREREQDLDPHLPGRHCK SCSHRNSTGNTDLMCSLCYLRYNNFIYKCAMGGGR"
AUTHORS	Engler, J.A. and van Bree, M.P.	1039. 1628
TITLE	The nucleotide sequence of the gene encoding protein Iva2 in human adenovirus type 7	/note="crossref Ad12 EPD07152"
JOURNAL	Gene 19 (1), 71-80 (1982)	/function="Elb promoter region (-499 to +100)"
MEDLINE	83054637	1542. 2033
REFERENCE	12 (bases 1 to 530)	/note="crossref SWISS-PROT:EIBS_ADE12, P04492"
AUTHORS	Shibata, H., Zheng, J.H., Koikeda, S., Masamune, Y. and Nakanishi, Y.	/codon_start=1
TITLE	Cis- and trans-acting factors for transcription of the adenovirus 12 ElA gene	/product="Elb protein, small T-antigen"
JOURNAL	Biochim. Biophys. Acta 1007 (2), 194-191 (1989)	/protein_id="CAA51878.1"
MEDLINE	89150250	/db_xref="GI:313363"
REFERENCE	13 (bases 1 to 34125)	/db_xref="SWISS-PROT:P04492"
AUTHORS	Juttermann, R., Weyer, U. and Doerfler, W.	/translation="MELETVLQSFQSVRQLLOVTSKNTSGFWYLFGLSLSKVNVK EDYREEFENLADCPGLLADLCLYLFQKVRVRLSDSSVGRVTSVAFIAFLATILDK WSKSHLSWDYMLDYNMQLRAWLKRRCVCIYSLARPLTMPLTLOEKEERPNAY VEK"
TITLE	Defect of adenovirus type 12 replication in hamster cells: absence of transcription of viral virus-associated and L1 RNAs	1740
JOURNAL	J. Virol. 63 (8), 3535-3540 (1989)	/citation=[9]
MEDLINE	89311650	/replace="t"
REFERENCE	14 (bases 1 to 34125)	/citation=[9]
AUTHORS	Kimura, T., Sawada, Y., Shinawawa, M., Shimizu, Y., Shiroki, K., Shimoto, H., Sugisaki, H., Takanami, M., Uemizu, Y. and Fujinaga, K.	/replace="g"
TITLE	Nucleotide sequence of the transforming early region Elb of adenovirus type 12 DNA: structure and gene organization, and comparison with those of adenovirus type 5 DNA	1847. 3295
JOURNAL	Nucleic Acids Res. 9 (23), 6571-6589 (1981)	/note="crossref SWISS-PROT:EIBL_ADE12, P04491"
MEDLINE	82105565	/codon_start=1
REFERENCE	15 (bases 1 to 34125)	/product="Elb protein, large T-antigen"
AUTHORS	Kimura, T., Sawada, Y., Shinawawa, M., Shimizu, Y., Shiroki, K., Shimoto, H., Sugisaki, H., Takanami, M., Uemizu, Y. and Fujinaga, K.	/protein_id="CAA51879.1"
TITLE	Nucleotide sequence of the transforming early region Elb of adenovirus type 12 DNA: structure and gene organization, and comparison with those of adenovirus type 5 DNA	/db_xref="GI:313364"
JOURNAL	Nucleic Acids Res. 9 (23), 6571-6589 (1981)	/db_xref="SWISS-PROT:P04491"
MEDLINE	82105565	/translation="MERIPPELGLHAGLHVNAAVEGMAEEGLHLLAGAFDHAATA DVARGEGAEPCGGVEVMEQVQEGHVDGEGPCADDDRQKKEKSLKEAAVLS RTVNLMSRPLEVYVWQELQDEPQRGMHLQYKYSFQKTHLWEPWEDMECAIKAF AKLARPCDSYRITKTITVTITTSATLNGAIVEDTSDRAVAFRCRMOGMPGVVGLDG ITFINRVFAGDKFGIMFEANTCLVLHGVIFLNFNSICVESWKNVSRGCIFYGCKWG
REMARK	(sites)	
FEATURES	Location/Qualifiers	

```
LVGRPKSLVKKLFEKCVLALIVEGDAHHRHNAASENACFVLLKGMALIKHNWCG
VSDQTRFVTCADNCHTAKTHVHSRHCWPCVDHNMFRCTIHLGRGMRPS
QCNFHSNIMLEPEVFSRVCLNGVFDLSVLECKVIRYNDTRHRRCRCEGSSHLELR
PVLVNLBELRSDHLTSLCLRTDIESSDDN"
2071..2073
/citation=[9]
/replace="gcg"
2417..2418
/citation=[9]
/replace="ga"
2479..2741
/citation=[9]
/replace="atactgaat"
3374..3808
/note="crossref SWISS-PROT:HEX9_ADE12, P03284"
/codon_start=1
/product="hexon associated protein (Protein IX)"
/db_xref="GI:313365"
/db_xref="GI:313365"
/translation="MNQTTQNNALFDGGVFSPLTSRLPYWAGVRQNVVGVSTVDGRP
VAPANSSTLYATIGPSPPLDTAAASAAASAASTARSMAADFSTYNHLASNAVTRAVR
EDILTMALKEITLTAQLELSQKVEELADATHTPAQPTQ"
complement(3844..5202)
/note="homologue to maturation protein Ad5, Ad2, Ad7;
Query Match 31.18; Score 275.4; DB 16; Length 34125;
Best Local Similarity 59.8; Pred. No. 3.7e-69;
Matches 481; Conservative 0; Mismatches 321; Indels 3; Gaps 1;
QY 68 ctccgtacagtaggagtgctaccctctttgagacagaaacccgcgtac---catac 124
D 32277 CACCATACACAACTACCACTACCGCCCTGTGAGGAACAGTCAAAAGCTACCTGAGTA 32218
QY 125 tggaggatccctcgtcgcgcgaatgtacacatttgacacatgcacaaacgtgagttacg 184
D 32217 CTTCGGAATTTCTTTATGCGCTGAGTGAATAGTCTGAGTACATAATGTAAGTGAGG 32158
QY 185 tgcaggtctccctcagctgaggtattacgtgattcaggaatgaggtgttccctggg 244
D 32157 TAAGAGCATCTCTCATGTGTAGGTTTACAGTGTGAGGAATGCCAATACCGTGGG 32098
QY 245 atatggttctaagcgggagaggtctgtaacctcgtgaggaagtgtatgcagtgctgctgt 304
D 32097 ATATGATTTCACTGATATGATGATGTTATTTTGAATAAATACATGATGATGATGT 32038
QY 305 gtgtgcacacattatcatatcacagacatgatcatcattgacaggttccgtggtc 364
D 32037 GTTGTGCCACTATAATGTTGAAGTTACTCAATATTATACATGCTATGATGCGGTGCTTA 31978
QY 365 tccactgtcattgttccagtcctcgggttccctcagctgagtgatagccgagcaggtttgg 424
D 31977 TTTATTTGTCATGCCAGCTCCGGGTTCTACATGATGATGATGATGATGATGATGATG 31918
QY 425 ccaagctggtttagagtggtggtgagtggtgagtggtgagtggtgagtggtgagtggtg 484
D 31917 GAGCGTGGTGAATGAGCGGTATATGCGGCTTAAATTAACAAAGGTGTTTTCGTATC 31858
QY 485 gggagtggtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 544
D 31857 GGGAGGTGTTTAACCAATTAATGCTTAAGAGGTGATGATGATGATGATGATGATGATG 31798
QY 545 ggggtgcacattatcattcctcgtcgtggtgagtgagtgagtgagtgagtgagtgagtgag 604
D 31797 GAGGTGCGCAATTAATTTACTTTAAATTAATGATGATGATGATGATGATGATGATGATG 31738
QY 605 cgcctatgagcttgatgacagcgttcgactgagtgagtgagtgagtgagtgagtgagtgag 664
D 31737 AAAAGATGATGTTTGAGAGGAGCGCTTAATTAATGGAATTTAAATTAACATGTTAGTGC 31678
QY 665 tgtgtcagtgattgctgatttaagtgagatcaggtgagtgagtgagtgagtgagtgagtgag 724
D 31677 TGTGTTGTGATTATGTAAGAGCTTAAGTGAGATGATGATGATGATGATGATGATGATGATG 31618
```

```
QY 725 caaggcgccttatgctgcggcggtgcgaatcatcgtcagagaccactgccaatgtgt 784
D 31617 CCAGACTGCTAATGTTAAAGTTGTTCAAGTAATCTGGAACACTGTGTCGCCCTCTAA 31558
QY 785 attcctgcagagcagcggcggtgcgagcagcagcttatttcgctgcgagcaccacc 844
D 31557 AACATAGTCGGCATGAACGTTATCTGTCAGCAACTGCTAAAGGTTTAATTATGATCATC 31498
QY 845 gccctatcctgatgcagcagattatga 869
D 31497 GACCAATTTTATTTGGAGATTATAA 31473
```

```
RESULT 8
ADERE4/c
LOCUS
DEFINITION Adenovirus type 12 DNA for early region E4. VRL 08-OCT-1999
ACCESSION X51800
VERSION X51800.1 GI:58543
KEYWORDS early region; early region E4.
SOURCE Human adenovirus type 12.
ORGANISM Human adenovirus type 12.
REFERENCE 1 (bases 1 to 3498)
AUTHORS Hogenkamp,T. and Esche,H.
TITLE Nucleotide sequence of the right 10% of adenovirus type 12 DNA
JOURNAL Nucleic Acids Res 18 (10), 3065-3066 (1990)
MEDLINE 90272430
REFERENCE 2 (bases 1 to 3498)
AUTHORS Hogenkamp,T.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-1990) Hogenkamp T., Institute of Molecular
Biology (Cancer Research), University of Essex, Pufelandstr 55,
4300 Essen 1, F R G
COMMENT See <V00043> for overlapping sequence.
FEATURES
source
1..3498
/organism="Human adenovirus type 12"
/strain="Hule"
/db_xref="taxon:28282"
complement(816..1691)
/codon_start=1
/product="early E4 34 KD protein"
/protein_id="CAB57854.1"
/db_xref="GI:6018382"
/translation="MQRDRRYRLAPYKYLPPCEQSKATLSTSENSLWPECNSL
TLNHEVRGIPSCVGTFLQEMPFPDMILTDEMFILKTYMSVCMCCATINVEVTO
LLHGERWLIHCQRPGSLQMSAGMLGRWFKMAYGALINKKCFWYREVNVHMP
KEYMYVGTFTVGRHLIYFKIMDGHAWLALEKVSFGSAFNYGILNNMLVLCDDYCK
DLSEIRMCWRPRTRLMLKVQVIAENTVRLKHSRHERYRQQLKGLIMHRAILF
GYNORENPWAADGH"
complement(1624..1986)
/codon_start=1
/product="early E4 13 KD protein"
/protein_id="CAB57853.1"
/db_xref="GI:6018361"
/translation="MPLPCIPPPVSRDTRACIAWLGLAHASCVDTLRFIKHDLKIT
PRAEVLASLREWLFPALTELRQRCQKRGALTSGRTWFCFFKEDARKSVVYDAAR
QTVSLQIGTIQVPTAL"
complement(1993..2343)
/codon_start=1
/product="early E4 11 KD protein"
/protein_id="CAB57852.1"
/db_xref="GI:6018360"
/translation="MKYCLRMVAGALTELFNINGLNLQNCVOIIQOWNENYLGMY
QSGSLMIEQFHDNALFLFTEIRAVALLAEVVEHLENRLQFLDLAVIFHOHSGDGRCH
LRDLRIQLIADRLD"
BASE COUNT 1131 a 864 c 620 g 883 t
ORIGIN
```

us-09-214-478-1.rge

Mon Jun 26 09:10:26 2000

Nucleotide sequence of the adenovirus type 40 inverted terminal repeat: Close relation to that of adenovirus type 5

Virology 156, 414-416 (1987)

2 (sites)

van Loon, A.E., Ligtenberg, M., Reemst, A.M., Sussenbach, J.S. and Rozijn, T.H.

Structure and organization of the left-terminal DNA regions of fastidious adenovirus types 40 and 41

Gene 58, 109-126 (1987)

3 (sites)

Ishino, M.

Analysis of structure and function of human adenovirus type 40 leftmost 1.85 kb region including transforming E1A gene

Sapporo Igaku Zasshi 57, 59-66 (1988)

4 (sites)

Ishino, M., Ohashi, Y., Emoto, T., Sawada, Y. and Fujinaga, K.

Characterization of Adenovirus type 40 E1 region

Virology 165, 95-102 (1988)

89265890

5 (sites)

Vos, H.L., van der Lee, F.M., Reemst, A.M., van Loon, A.E. and Sussenbach, J.S.

The genes encoding the DNA binding protein and the 23K protease of adenovirus types 40 and 41

Virology 163, 1-10 (1988)

88160034

6 (sites)

Kidd, A.H. and Erasmus, M.J.

Sequence characterization of the adenovirus 40 fiber gene

Virology 172, 134-144 (1989)

89370295

7 (sites)

Toogood, C.I., Murali, R., Burnett, R.M. and Hay, R.T.

The adenovirus type 40 hexon: sequence, predicted structure and relationship to other adenovirus hexons

J. Gen. Virol. 70, 3203-3214 (1989)

9011698

8 (bases 1 to 34214)

Davidson, A.J., Telford, E.A., Watson, M.S., McBride, K. and Mautner, V.

The DNA sequence of adenovirus type 40

J. Mol. Biol. 234 (4), 1308-1316 (1993)

94087748

9 (sites)

Pieniazek, N.J., Slenenda, S.B., Pieniazek, D. and Luftig, R.B.

VA RNA region of human enteric adenovirus type 40 shows homology to the VA region of simian adenovirus type 7

Unpublished (1992)

Location/Qualifiers

1..34214

/organism="Human adenovirus type 40"

/isolate="Dugan"

/db_xref="taxon:28284"

/tissue_lib="ATCC VR-931"

1..163

/rpt_type="inverted"

387..392

/notes="for E1A"

418..>418

/product="E1A"

/evidence="experimental"

join(486..1011,1088..1311)

/standard_name="E1A 249R"

/codon_start=1

/protein_id="AAC13949.1"

/db_xref="GI:303971"

/translation="MRMLPDFFTGNWDMFQGLLETVYVDFPEPEASEEWSLHDLF DVEVDGFEDANQAVDGNFPELLSEAFSAAGSGDGVGEELPVLVDLKDVCYEDG LPSPDPTDEAEAEAEAAFTYNNENELVLDCEP&RGCRACDFHRTSGNPEA MCALCYMLTGHCYISIDAESESGSDDFPHPTATPHGIVRTIPCRVSCR RRAVECTIDLEDPTDEPLNLSLRPKCS"

join(486..669,1088..1311)

Query Match 30.9%; Score 273.8; DB 16; Length 3498;

Best Local Similarity 59.6%; Pred. No. 8.9e-69;

Matches 480; Conservative 0; Mismatches 322; Indels 3; Gaps 1;

QY 68 ctccgtacagtgaggatcgtctacatctcttttgagacagaacccgcgcgtac---catac 124

DB 1657 CACCATACACAAAGTACCACCTACCGCCCTGTGAGGACAGCTCAAAAGCTACGTTGAGTA 1598

QY 125 tggagatcatcgcgtcgtccgcgaatgtacacatttgacatttgacacacgtgagttacg 184

DB 1597 CTTCGGAAATCTTTATGCGCTGAGTGTATAGTCTGACTTTACATATATGAGTGGG 1538

QY 185 tgcgaggtctccctcagtgaggattacgctgattcaggaatgggtgttccctcgtgg 244

DB 1537 TAAGAGGCATCTCTTATGCTGTAGTGTATGAGTGTGCGAGNAAGGCCAATACCGTGGG 1478

QY 245 atatggtctaaacgaggagagctgttaacccctgaggagtgatgtacacgtcctggttc 304

DB 1477 ATATGATCTTAACGATTAATGAGATGTTTATTTGAAAAAATACATGAGTGTATGTATG 1418

QY 305 gttgtcccaacattgatatacatcagcagcagatgatcattcattcagcagctcgtgctc 364

DB 1417 GTTGTGCCACTATAAATGTTGAAGTTACTCAATATTACATGTCATGAGCGGTGCTTA 1358

QY 365 tccactgtcattgtccagtcctccggttccctgcagtgatagccggtggcaggttttgg 424

DB 1357 TCCATGTCATGCGCAGCGTCGGGTTCACATACAGTGTATGTGAGTGGGATGCTTTTG 1298

QY 425 ccagctgtttagatgtgtgtgagtcggtccatgttttccatcagagtttatgtgtacc 484

DB 1297 GAGCGTGTGTAATGATGCGCTTATGCGCTTAAATTAACAAAGTGTGTTTGTGTATC 1238

QY 485 gggaggtgtgaattacacacatgccaaagaggttaattttatgtccagcgtgtttatga 544

DB 1237 GGGAGGTGTTTAACTTAATGCTTAAGAGGTGATGTATGTGGAGACGACCTTTGTTA 1178

QY 545 ggggtgcgcacttaattacatcctcgcgtctgtgtgtatgagcagctggttctgtgtcc 604

DB 1177 GAGTGGCCATTAATTTACTTTAAATATATATGATGCGCATGCTGTTAGCGTAG 1118

QY 605 ccgcatgagcttgatgacacgcctctgacgtggtgatttggaacatttgggtgc 664

DB 1117 AAAAGTATGTTTGAATGAGCGCTTTAATATGGAATTTAAATTAACATGTTAGTGC 1058

QY 665 tgtgtgcagttactgtctgatttaagtgcagatcaggtggtgctgctgctccgagga 724

DB 1057 TGTGTTGTGATTATTGTAAGAGCTTAAGTGAGATAGCGCATGCGCTGTGGCCCTCGTA 998

QY 725 caaggccttatcgtcggcggtgctggaatcctcgtcagagacacactgcatgttgt 784

DB 997 CCAGACGCTTAATGTTAAAGTTGTTCAAGTATTGCTGAAACACTGTTCCGCTCTAA 938

QY 785 attcctcagcagcagcagcggcggtgagcaggttattcgcgcgtgctgcagcaccacc 844

DB 937 AACATAGTCGCATGAACGTTATGTCAGCACTGCTAAAGGGTTTAATATGATCATC 878

QY 845 gccctatcctgacacgattatga 869

DB 877 GAGCAATTTTATTTGGAGATTATTA 853

RESULT 9

ADRGENOME/c ADRGENOME 34214 bp DNA VRL 17-APR-1998

LOCUS Human adenovirus type 40, complete genome.

DEFINITION Human adenovirus type 40, complete genome.

ACCESSION L19443

VERSION L19443.1

GI:303969

KEYWORDS Human adenovirus type 40.

SOURCE Human adenovirus type 40

ORGANISM Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.

REFERENCE 1. (sites)

AUTHORS Ishino, M., Sawada, Y., Yaegashi, T., Demura, M. and Fujinaga, K.


```

/standard_name="E1A 135R"
/codon_start=1
/protein_id="AAC13948.1"
/db_xref="GI:303970"
/translation="MRMLPDTFGWDMFQGLLETKYVDFPFPESEASEMSLHDLF
DVEVDGEEDANQAEVDGPIFASLNHGHSPFQIWRNLDFSPGRTVSGIAFCIFLDQ
CRPRAVEICIELLEDPTDEPLNLSLRPCS"
<486..1011
/standard_name="E1A 249R"
/number=1
<486..927
/standard_name="E1A 221R"
/number=1
<486..669
/standard_name="E1A 135R"
/number=1
Join(486..927,1088..1311)
/standard_name="E1A 221R"
/codon_start=1
/protein_id="AAC13950.1"
/db_xref="GI:303972"
/translation="MRMLPDTFGWDMFQGLLETKYVDFPFPESEASEMSLHDLF
DVEVDGEEDANQAEVDGMPFERLLSAAESGSDGSGVEILLPLVDLKKCYEDG
LPSPDTEATEAEAAEMPTVYNNENELVDCPENRCRCRACPIADAGESESG
SPEDTPHPLTATPPHGIIVTICRVSCRRRAVSCIEDLLEEDPTDEPLNLSLRP
KCS"
<603..603
/note="E1A-E1B cotranscript"
/number=1
/evidence=experimental
1088..>1311
/standard_name="E1A 221R"
/number=2
1088..>1311
/standard_name="E1A 249R"
/number=2
1088..>1311
/standard_name="E1A 135R"
/number=2
1321..1326
/note="for E1A"
1373..1378
/note="for E1B"
1403..>1403
/product="E1B"
/evidence=experimental
1407..>1407
/note="E1A-E1B cotranscript"
/number=2
/evidence=experimental
1412..>1412
/product="E1B"
/evidence=experimental
<1414..3151
/standard_name="E1B 19K"
/number=1
1414..1917
/standard_name="E1B 19K"
/codon_start=1
/evidence=experimental
/protein_id="AAC13957.1"
/db_xref="GI:303979"
/translation="MEIWSLQSYQNLRLLEASARTSSCWRIILFGSTLTNVIYRAK
EYSSRFDLLSHNPGIFASLNHGHSPFQIWRNLDFSPGRTVSGIAFCIFLDQ
LDFSTEL"
<1719..3151
/standard_name="E1B 55K"
/number=1
1719..3149
/standard_name="E1B 55K"
/codon_start=1
/evidence=experimental

```


388	ggtccctcgagtgatagccggcagggttttggccagctggttagcatggtggtg	447
32385	GGGTCCTTCAATGTAAAGCGGGGGTGAGTGTAGCAAATTGGTTTAAAATGATGATT	32326
448	gatggcccattgtttaaccagaggtttatatggtaccgggaggtggaatacaacatg	507
32325	TATGGTTCACGTATATACTTAAGATCCCTGTGTATCGAGAAATGTAATTTGGCATG	32266
508	cctaaagaggtaagtatttatgtccagcgtgttttatgaggggtcgccactaatctaccctg	567
32265	CCCAGGAAGTGTATTATTATGGGAGCGCTTTGTGAGAGGTAGGCATTTTATATATATT	32206
568	gcctgtggtatgatgccacgtgggtctctgtggtccccgcctagcgttgtgatacacgc	627
32205	CCTCCTGCTGTATGATGGCACCGGCTATAGTGTGGAGCGCATGCTGTTGGTGGTAGT	32146
628	gcttcgacctggtggattttgaacaataattgtggtctgtctgagttactgtctgat	687
32145	TTTTTTAGTTATGAATTAATGAATATGGTGTGGTGTGGCTGTACCTATTGTAAAAAC	32086
688	ttaatgagatcagggctgcctgctgtgcccgaggaagaaggcgccttatgctgcggcg	747
32085	TGGATGAGCTTCTGCTCGGTGTGCGCTCGTGGACGGCGGCTGCTGGCCAAGCGG	32026
748	gtcggaatcctgctgaggagaccactgccaattgtttattcttcgaggaacgagcgcg	807
32025	GTCMAAGTGTGGGAAGCTTACCCTTAGGTCATCATGAGGAGCTTGGTGAACCCCGG	31965
808	cggcagcaatttatcggcgcctgctgcagcacacacgcgccttatcctgatgcacgattat	867
31965	AGACAGGGATTACTCGGGGCTTGATGAGCGGTACTGCCCGTTTACGTTGGCAGACTAC	31906
868	gac 870	
31905	AAC 31903	
RESULT 10		
AB026117/c		
OCUS		
DEFINITION	Porcine adenovirus 3 DNA, complete genome.	" 22-APR-1999
ACCESSION	AB026117	
VERSION	AB026117.1 GI:4630864	
KEYWORDS		
SOURCE	Porcine adenovirus 3 DNA.	
ORGANISM	Porcine adenovirus 3	
REFERENCE	Viruses; GSDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.	
AUTHORS	Larocque,D.	
TITLE	porcine adenovirus serotype 3, complete genome	
JOURNAL	Unpublished (1999)	
REFERENCE	2 (bases 1 to 34094)	
AUTHORS	Larocque,D., Malenfant,F., Massie,B. and Dea,S.	
TITLE	Direct Submission	
JOURNAL	Submitted (14-APR-1999) to the DDBJ/EMBL/GenBank databases. Daniel Larocque, INRS-Institut Armand-Frappier, Biotechnology; 531 boul. des Prairies, Laval, Quebec H7N 4Z3, Canada (E-mail:daniel.larocque@videotron.ca, Tel:(450) 687 5010) Location/Qualifiers	
FEATURES		
source	1. .34094 /organism="Porcine adenovirus 3" /db_xref="taxon:35265"	
repeat_region	/note="porcine adenovirus 3 strain IAF is a high titer clone, derived from strain 6618 and adapted for cell line PKA (pig kidney, porcine circovirus free)."	
TATA_signal	line PKA (pig kidney, porcine circovirus free).	
CDS	1..144 449..452 /note="Possible promoter region of E1a genes"	
	533..1222 /note="Description: E1a; homologous to E1a 289R of human adenovirus 2 and 5; putative early genes transactivator, induction of host cells DNA synthesis and activate p53 downstream apoptosis cascade"	

[illegible]

RESULT 11
PAD237815/c 34094 bp DNA VRL 04-MAY-1999
LOCUS Porcine adenovirus serotype 3 complete genome, strain IAF.
DEFINITION
ACCESSION AJ237815
VERSION AJ237815.1 GI:4678602
KEYWORDS 100 kDa protein; 202R; 214R; 229R; 474R; 52 kDa protein; COOH
region; DBP; E1A; E1B; fiber protein; hexon; IIIa; penton protein;
pIX; proteinase; pTP; pVII; pVIII.
SOURCE
ORGANISM Porcine adenovirus 3
Virus; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
REFERENCE 1 (bases 1 to 34094)
AUTHORS Larocque, D., Malenfant, F., Massie, B. and Dea, S.
TITLE porcine adenovirus serotype 3, complete genome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 34094)
AUTHORS Larocque, D.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-1999) Larocque D., INRS-Institut Armand-Frappier,
531 boul. des Prairies, Laval, Quebec H7N 4Z3, Canada
FEATURES
source
1. 34094
/organism="Porcine adenovirus 3"
/strain="IAF"
/db_xref="taxon:35265"
/note="high titer clone derived from strain 6618 and
adapted for cell line PKA (pig kidney, porcine circovirus
free)"
repeat_region
1. 144
/rpt_type="INVERTED"
/rpt_type="TERMINAL"
TATA_signal
449..452
/gene="E1A"
/note="possible promoter region of E1A genes"
gene
449..1291
/gene="E1A"
CDS
533..1222
/gene="E1A"
/function="putative early genes transactivator, induction
of host cells DNA synthesis and activation of p53
dependent apoptosis cascade"
/codon_start=1
/product="E1A 229R"
/protein_id="CAB41021.1"
/db_xref="GI:4678604"
/translation="MANRLHLDWDGNPEVVPVLEWDPVDLRDPSGDGFCPCWESL
VGLPDELSDVDEVIIVTEGESDSSGSAAGSGSQGVFEMDPPEGDSNEEDI
SAVAEVLSELADVFEDPLAPPSPFLDPCPEVGVNCRSCDIHRFHSKDPNLKSLC
YMRHAFVYGCFTFTVGLCGKKGKSKSHVLPFICLFRSCFSSTSQVGFPTW
RPAQREVELL"
join(533..1043,1141..1274)
/gene="E1A"
/codon_start=1
/product="E1A 214R"
/protein_id="CAB41020.1"
/db_xref="GI:4678603"
/translation="MANRLHLDWDGNPEVVPVLEWDPVDLRDPSGDGFCPCWESL
VDGLPDELSDVDEVIIVTEGESDSSGSAAGSGSQGVFEMDPPEGDSNEEDI
SAVAEVLSELADVFEDPLAPPSPFLDPCPEVGVNCRSCDIHRFHSKDPNLKSLC
YMRHAFVYGCFTFTVGLCGKKGKSKSHVLPFICLFRSCFSSTSQVGFPTW
RPAQREVELL"
CDS
1286..1291
/gene="E1A"
/function="putative inhibitor of apoptosis (Bcl-2
homologue)"
/codon_start=1
/product="E1B 202R"
/protein_id="CAB41022.1"
/db_xref="GI:4678605"
/translation="MDYQLLAKLTNNVLRKVIVQSGNCPWKKIFSDRFKIVVAEA
RRQYQGLIEIFVEGERGFPEFLRGLGLYEAVLKLDELSTLGRTVASVALVCFIFE

KLOHSGWTDEGLLSLVLPLCLSLLEARMMAEQVROGLCIIRMPISAERMLLP6SGSG
SGSGAGMRDQVVPKPRRQEEDEEDGMEASGRRLGPDIV"
1829..3253
/function="putative inhibitor of apoptosis by p53
dependent mechanism"
/codon_start=1
/product="E1B 474R"
/protein_id="CAB41023.1"
/db_xref="GI:4678606"
/translation="MFPAGGANDGAGAGAVHODAEERGAGDAVAOWVIRQWGRD.
AGPGAQAPAGAGRGGGGWDGSEARAGASGLDRRRPGGAGGEGSGEAGGSSM
VSYQVLSLEYLSEPLEMHRYSEFQIPIMLOPGDDLGEMIAQHAQKVLQGTVVELR
RPTIRSMCYIIGNGAKIRGNTEYINIEPRNHCMSIAGMSVTITDVVDELPA
RGLLIANTHFILHGCNFTLGSVITANAGGVRCGYFCACIKALDRHRLMTVNE
NFEKCVAVVSAGRCRIKNSLSLSTFCFLHMSYTKIVGNSIMSPTYSDDPTVLLV
CQSGMPLSTVHIAPSRPLYPFERNVLLRSTMFVGRGLSGFSRCSYSLSLV
VDESGRGLSVTCCEDTCTCMYKLLQCTADEMETDTSQYACLCGDNHPVQVRQMK
VTDALRAPRSLVSCNMGFEFSDDD"
3394..3993
/function="hexon-associated protein"
/codon_start=1
/product="pIX"
/protein_id="CAB41024.1"
/db_xref="GI:4678607"
/translation="MSSEGEIRTCFISARLPWSAGVROGVAGTVNGGVGAPAGQSV
LAYSREFVQOQOQPGTAATGVSFRAVFPSPVDSAEVGMRLAELRQQLREVEVE
IQLRATASEAEEEEIEIVDEEAPAGAGANTEEEEDEMLTMTVGDPEPAGVEA
QPPPTTPSPDPAVPATTTTPKRLSYGASKRSGPCAEEN"
4085..4090
/gene="E1B"
/gene="E1B"
4085..4090
/gene="E1B"
complement(5962..5965)
6009..6015
/gene="major late promoter (MLP)"
6009..6015
/gene="major late promoter (MLP)"
complement(8555..>10396)
/function="essential for DNA replication"
/codon_start=1
/product="pTP (COOH region)"
/protein_id="CAB41025.1"
/db_xref="GI:4678608"
/translation="VELFRPVNINWRAHOWTRASVAVGIAMSRFIYRHRMLMN
LSPROATEHPPLXLYPPHLLGYOHLRLCNDYIFTRAYSRLRTEVTPHYQVY
NCPWATGTYINTGAYHRFLDDDFSDLSQMOQAIAERVAVDALLRLPGYGT
HDDARDVPVHLLQELQKIDRCQOQEWANGADRILQAGRKDLTLRLQRLKLTAY
FHLVSRHLRPLTHALNLPCCWLDADFVERFAGSEQVREAEALQOQVPTRELT
RCIIISALSLPCLSGTALRGAPFELRPGGGAATVETMRMRKGEVIERVDRPLRRR
PPAPVPVAAAEAEDEEMPELEGEAEAAVAGPSTPPSPFEEERQVAAEALIA
LEELTVSAQHEFFRFAPEFVVRMORLDGIDNELTLRRWVINFFLCEHIATTLNT
LHALLQAPFNRLVALNAQVVMRADQDQGVYSRVNGENALVAMNRIITVLI
AATVRAHGHDLDDELEQFMQDIARNSGDVSEILRQVATHDADIDSVLSFRFY
TGPVFSFGHQIQINRRVIAHASRLRDHLPLPELNAPVPLAQLPA"
10629..11666
/codon_start=1
/product="52 kDa protein"
/protein_id="CAB41026.1"
/db_xref="GI:4678609"
/translation="MAPPAGALCAEATRSRPSLTPRREEGEARLGAIAAPER
HPVQLKREAMEAYVPQNARFRPGESEERDLFRFRAGRMQDRERVLPQDFEG
RVFEAGVSAARAHMSAAQAIQETVREEVNFQKTFNNVTVTLVSRDEVTMGLMHL
WDFEAFAPLQHRSAALTAQLLQACHCRDEGMVKEALLSLGASRWLVLDLYNLQTI
VQERSMSLSEKVAALNYSVATLAKHYAKISTIFYMRAVVKLLVLADLGMTRNEKLE
RVYSTSRRELNDLELMFLRRALAGEBDELEEDLEEELEERGGVRSPPGTAA
RENAVADACER"
11718..13586
/codon_start=1
/product="IIIA"
/protein_id="CAB41027.1"
/db_xref="GI:4678610"
/translation="MAASAGAAAGVSAASLAPERATRMOALPSIDPEWQALRRI
MALTADUSRRFASQPLANRIGAILLEAVVPRTNPTHEKVLTVVYVVALLETSAIRPDEAG

[illegible]


```

AGATSTIIVNPKRYKGVAVVQVQATLSISAVLNGOCVEIKYHSNLSASLTNLCNTNL
YDLPACLNRPITAHNLPALIEBAAPSLICYVQRGVRAVEFAEVLPLLSFKLFLV
KGGVFLIKDISOMOKCEFCGFFKHTVTCALARRPFIHHVAHSAADWFLISFTPI
GAPPNTELFIVDYVEITYWGHGKGLFQVFLQGLDGHVNVKQPOLQEKATLIG
WNGEODTLYCITPEKRAIGVFKFRDTLQHIAASLWSHLVICONPOLQEKATLIG
LESPEELTPDOLKFKFGLKGNPRFIEVAVGNHNTGDEILLAAQVVSUPEPVEI
CRNMPRAGRLNFNDITYSLPNFSPYPAKSEYHEWQGVLAADLSQIKFAMVRDTPS
LTHSLKNAKAYDILQOTLDYCALDVQVTAQLVNLKIESQYIFKNSVNLPEFTSNFV
ORPTISNSHAFKQILYKAEKPNTHSLTILMAWSENMEIEVRLSIRGRCYPTIG
VLOPFVYDIOGMYASALYTHPPAGSPNPLERAKAKAEQKMLNLSIYFQDKOL
LPGLFTTTPPEKWLAEVYVNLNLSAKEADKSNQTMRSIAKLNSLALYGSFAFK
LDNKKVFSODIENSAKAGVAVKSSYIETDNLCAETMEPEFVAVPVNSDVR
QLAPSCSEDPKTOPLAEAPFMHNSMTSYHYKPIFIDAEDDDCLHLEKSTPLI
ANNRYFSQASFLWATRAFYSEWQFLYENDAGTLENVRLKSYGDTDSLFTEG
YKMEERKKRLKNGKLVDFPSNPETLWVECTOCEKGSADISESYLAPLKY
ALMDTTPCHVHGKGLRAKHATSTLSYDLVLCACYADMQGSDVFKTSMRLRT
LTSVQAHVQFVTETTLTRKLRPKDKTLHALDMHRLIPYSRHPNPRNTETWMEL
OWMT"
complement(join(7941..9764,12575..12583))
/function="DNA replication"
/note="pnp"
/codon_start=1
/product="terminal protein precursor"
/protein_id="AAB38717.1"
/db_xref="GI:1732272"
/translation="MSLNALDCARLTGQTYTVEVRPLNINFRVREYTRASTSVG
LAWMSKYIQVHRLMLNLSPREPATEGPLFWPPHLLVGYQYLVRCNDYVDFDR
SVRLKYLHLPQKLNWTVANCSTYINTGAYHRFIDFENFAETLAOVOAVLAE
RVADLALIRPAGTYTINMAGDQVPEGLLDQHYKNLSQCONAEWGLDMRQQA
GKDLVILTRKLTAFNPLSPRNPTMLSCDCLWDLAFKEDTPELSQFT
IQSLPSQVTSIIISALSRLPAPCTPLSGAEPFLPRENGPRAVEEMRRRGEVIER
FIDLPVRRRRRPPPPVSELSPEAEAPPPAPSPRASFEFEFVDIVAEIRL
LOELTYSARNEQFFFAVDFYEVQIORMLGNELTIRRWVDFVFAEIVATLRL
LHNLRLYPPCSRWVDELAQVYMRARDHQGVYSKRVNEMGENAFSQVMARYSGDL
AATVERAGLEEEEMQPMADIAHENGSDVSEILLROVAINDTVEVDSMELSPREKV
9814..10980
/function="viral assembly"
/codon_start=1
/product="52.55kDa protein"
/protein_id="AAB38718.1"
/db_xref="GI:1732273"
/translation="MHPVLROMKPLPASAKLEAESEGLARLOGGAPELHPVOMKQ
EAEAVIPTSNVFRNEGEAEGLHLYAESGRLLRNDHTSKRVIGERDPEKQNGI
SAAEHLKSAADVAYEHTVRAEVNFQTFNNVRLTIAREEVVIGLHLMDEVEAPL
ENPVSAQLTQLFLVQHCDEGLVRESLNIAPESRWLDLNLGLTIQVQERGLA
VGEKVAINISVITLKHARKIFNSVFPIDKEAKINTFTYMETVVKLVLSDDLGM
RNERIAVAGSARQRELNDRELMHRLRQALASNLPLEGEDSANSIRKKEWVGAGG
GVASARYPHLLDYEEENPDGVSFQHERGAQPHENGHAESAYSRRQLGRFY"
10853..12556
/function="structural protein"
/note="pili"
/codon_start=1
/product="pili protein precursor"
/protein_id="AAB38719.1"
/db_xref="GI:1732274"
/translation="MKRRRLMAPSHFNSMSALNPKMAAMQSOPTDADDGWAASIR
IMSTADGRHFKSQPPANRLDALLAEVPSKDPHEKVLITVALIENGAIARDEG
AGVYDALLHCKAYNSLNAQSNLERLAGVREAVQAVQRIATNGLSITANGSLARL
PANYRGQENTGVFSALKLVSEVSTEVQSGHYFLQSRNGTQVNTLNAFENL
KPLWGVQPMERLSTISALLTPRULLLVFFDVSISDSYSLGVLTLTYREALG
RNHLDERTLEVTEVRAGSNNINLQATLNLNRQKPKDYSITLPEERIVRF
VQVARSIMQENLSTEDVYTAANMESFYASNRDFINKLMDFPHRAAMADIFL
GAYNRPWLPFFGFFGVDFPERDYSYTWGDLSSLELTRODAMRFLDFKMDQDQRT
ESRSLSRVTTPASSRSYVMPGGFMGNKNNKNSLRIDGLAKLRWKTYKRE
SEARVSLPAVRPKQYRPSRPSISSDDSDGMSRDPFLKPEGSGNPFALRPLKRG
L"
12607..14040
/function="structural protein"
/note="viral component III"
/codon_start=1

```

CDS

```

Query Match
Best Local Similarity 5.68; Score 49.8; DB 16; Length 31323;
Matches 117; Conservative 0; Mismatches 112; Indels 0; Gaps 0
QY 356 cctgggtctccactgtcattgttccagtcctccggtccctgcagtgatagcgcggcgc 415
DB 29026 CGTGGCGAGTACACTGCTACTGCTTAAACAGCTTGCTTTACAAATGCGTAGCAAGC 28967
QY 416 aggttggccagctggttagatggtggtgagcgcctgtttatcagagattta 475
DB 28966 ATGNGTTCAGAAAGTATGAGGACTTTTATAAGGGTGGGCTATGATAAAGATATA 28907
QY 476 tatggtaccggggagggtgggaattacaacatgccaaagagagtgatgtttatccgcg 535
DB 28906 TGTGGGTACCGAGAGTTGTAAATTTCTAGCAGCCCTGATGAATTAACATATGCGGAGTA 28847
QY 536 tgtttatgaggggtgcacctaattcctcgcctgctgtggtgatggtg 584
DB 28846 TTATATTAGAACACTCACTACATATTATTTTAGACTGCTCTTTTITAG 28798

```

RESULT 15
CAUS5001/c
LOCUS

DEFINITION

CAUS5001 30288 bp DNA

Canine adenovirus type 1 E1a protein, small T-antigen, large protein, p11a, penton base protein, pVII protein, pV protein, Mu peptide precursor, pVI protein, viral protein II, endoproteinase, E2A-DBP, 100 kD protein, L4-22kD protein, L4-33 kD protein, pVII, virion component IV, U-exon, E4 orf5, E4 orf4, E4 orf3, E4 orf2, and E4 orf1 genes, complete cds.

ACCESSION

VERSION U55001.1 GI:1477654

KEYWORDS

SOURCE

ORGANISM

Canine adenovirus type 1.

Canine adenovirus type 1.

Viruses; GSDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.

REFERENCE

1 (bases 1 to 199)

AUTHORS

Sira, S., Abouhaider, M.G., Liu, Y.C. and Campbell, J.B.

TITLE

Multiple reiteration of a 40-bp nucleotide sequence in the inverted terminal repeat of the genome of a canine adenovirus

JOURNAL

Virolgy 159 (1), 76-83 (1987)

MEDLINE

87265479

REFERENCE

2 (bases 24577 to 27568)

AUTHORS

Draquelev, B.P., Sira, S., Abouhaider, M.G. and Campbell, J.B.

TITLE

Sequence analysis of putative E3 and fiber genomic regions of two strains of canine adenovirus type 1

JOURNAL

Virolgy 183 (1), 298-305 (1991)

MEDLINE

91272490

REFERENCE

3 (bases 1 to 30288)

AUTHORS

Campbell, J.B. and Zhao, Y.

TITLE

DNA sequence and genomic organization of canine adenovirus type 1

CDS

CDS

CDS

Ad2" /note="splice site tentative, by comparison with human
Ad2"
/codon_start=1
/product="iVa2 protein"
/protein_id="AAB05433.1"
/db_xref="GI:1477659"
/translation="MEEKAGSLHQLNQNEPSODPHQCADPCPALHSDRNLHKEAEA
MEGONPTCSRHESAPYIQSQSKSKQKQRYVGAAYDDILKNLWDRLQTLQOOSTEMPY
AGKPLKKNFTFEELSLQGSLLNDLIDQITQCCIRASKYLKNGEKSTLEMY
KQPIAFLVPTGCGSKQLNLMSAQLIVPTVEFTTQCCIRASKYLKNGEKSTLEMY
CEGNLAGENPTVPGSQIMPFKIQMSAQLTKDKNYDVTSPNNIPANASGPITAL
VNDECMDELGNNRIGIAKTFHAFSPKLDKFPKCKTSYVIVLHNNMPRDOQGNIAL
KIQAKVHMTSPKVHPQSOLNRFNIYTKGQPTASLLALKDITFHLRLNTFDWYINTE
PIDNCWHMWSLSPQEGILPMYLNLSQALQALERIHTLTDROKRTYIYHSKKK"
complement(join(4593..8033,12508..12516))
/function="DNA synthesis"
/note="splice site tentative, by comparison with human
Ad2"
/codon_start=1
/product="DNA polymerase"
/protein_id="AAB05434.1"
/db_xref="GI:1477660"
/translation="MSLVQSHGTSGLTEPPNSINQOESSGSPSLPAODATQOASASSAR
AGATPAINTSKRYGAVAAQRATLSAISLDNGQCQVIQYHSLASALNLCNTNLY
DLAPLCKNTIPNORLPIEBEASYSIIYYOQGTVRVFAEILSPFLKFLYK
HGKVLKLDSPMQRCBCEFSKATYTLTRRRDFVHHVASHSGDWMKEISFPIG
ANATERLIVYDVTWIGKFGKQLVPMFLPQLGDELHVNAAKLNATNWDTW
NSNEDTIVCINPEKRAIGVKFTFRTDQLQHFHNSHLVCLCNPKLMEAEALGL
ENEDITANOLKFKLGLOTPRFIBVVVGNHITGQFDELLAAQVVSRAEIPPVDFIS
RNFMRAGRLNFDITISLNPSPVSKDVHMGQGVLASDLQATQIKFVROTFTSL
TSLSKNAKASITSVSKGCCPYQAVNEFYMLGSYQDQAGFPDLKVKWQDOEYCFNK
DLWKKKGAGDIIOQLDICALDVQATVLNKLIDLSQYFIRLSIRGRCYPIYIG
DMLTKSSHAIFKQILYRAEKNAPHLNTLIMPASNEMIEYFIRLSIRGRCYPIYIG
LOGEIVFDICGMTASALTFFPAGSLNTPYERAVAKIAPHKQOEHKTIISFDEKLL
PGFTIDADPPAEFLLVPCSKGGKGLCTNEPLKEGYTTSIDVITLHNRGDDTL
LIPRTVTPPEWKLAREYVOLNINAKEKADKSKNQMTRSIKLLSNALXGSAFATL
DNKTSVEEYTPENHAGBPQWQNFNTSYRYKPTIDEADDDDFCLHLEKSTPLIT
NNRYSQASITLWATRAFVSQNSQYENDAGIPEKRIKLSVYGDTDSLFTTWEGY
RLMEKQGRKRLKNGNVLFDPNPPELLTWVECTKCKBKAGADAYSSSVYAPLAKYA
LKDITCPQCQVGTGKLRAGHATSTSYDLVKKAYADLQOQSDIFKIRMSLRILT
TSVOTHQVQPTVETTLTKRLRPWKDKTLHALDMNRLTYISKYPNPNNETTWMLQ
WMT"
complement(join(7889..9706,12508..12516))
/function="DNA replication"
/note="terminal protein precursor; splice site tentative,
by comparison with human Ad2"
/codon_start=1
/product="tpf"
/protein_id="AAB05435.1"
/db_xref="GI:1477661"
/translation="MSLNALDCARLTGOTPYVEVFRPIRNFIRNREYTRASTTSVG
LAWMSKRIYIHLRLMNLNPREPTEGWPLYPPHLLYGVQLVRVTCNDVDFTR
SYSLRYKIYIHLPLQOOLNLYVMANCSTINGAHRFTDFNFEETLAQVOQVLAIE
RYVADALIRNMGVNTMAGDVRVQEGLLQDHKLNLSQCNHAGLADNRIONIA
GNKDVITILKRLATFNFLVSRPNHTIILSPCCQLQCNHAGLADNRIONIA
IQSLPSQVTKGIISALSPGAPCTLSGGFAFPAPRRRSEEEVYDTEIVEARILQL
FIDRLPQRRRRRAPPMPMSSELSEPEVAEPAPRRRSEEEVYDTEIVEARILQL
ELITYARNQEFNFALNPEYIQRLEMLGNLTNELIRRWVYFVAEHVATLNTLH
HNLRLTPCQSRVDLELAQVNRADHEGOVYSRVNMENEMAFSOLMARVSGDLAA
TVERAGLGESEEMQFADATYAHENSQDSEILQVAINTEVDMSMELSPFKFTYK
PVYFSQNRQIQRNNRVVALASQLNRMQHRLPLAQHQEQVQLP"
9756..10925
/function="viroin assembly"
/codon_start=1
/product="52-55 KD protein"
/protein_id="AAB05436.1"
/db_xref="GI:1477662"
/translation="MHPVLRQMKPLPASVVKLAESEGLARLOGGAPEVHPVQMKQ
EAAEYIPASNVFRDEGEAEGRHLKYESGRLKRDHFSKRVLDERDEFEKSPENI
ENPASHALKADLTATYETVKAENVFTQPTNNRVIRLAIREEVYIGLHMLDVFEAFL
ENPASHALKADLTATYVQHCDEGQVRESLLINIAEPBSRWMLDLNLQITIVQVRSIA
VGEKVAALNYSVITLSKHARKYFNFSVEPDIKEAKINTFYMRTYVQIIVLSDSLGNY

	CDs	CDs	CDs
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			
46			
47			
48			
49			
50			
51			
52			
53			
54			
55			
56			
57			
58			
59			
60			
61			
62			
63			
64			
65			
66			
67			
68			
69			
70			
71			
72			
73			
74			
75			
76			
77			
78			
79			
80			
81			
82			
83			
84			
85			
86			
87			
88			
89			
90			
91			
92			
93			
94			
95			
96			
97			
98			
99			
100			

```

JOURNAL      Unpublished
REFERENCE    4 (bases 1 to 30288)
AUTHORS      Campbell,J.B.
TITLE        Direct Submission
JOURNAL      Submitted (14-APR-1996) James B. Campbell, University of Toronto,
              Microbiology, 150 College Street, Toronto, Ontario, Canada, M5S 1A8
COMMENT      Nomenclature of proteins has largely followed accepted terminology
              of the human adenoviruses.
FEATURES     Location/Qualifiers
              1..30288
                /organism="Canine adenovirus type 1"
                /strain="CIL"
                /db_xref="taxon:10512"
                /lab_host="dog kidney"
              1..199
                /note="left ITR"
              54..93
                /note="40-base pair repeat not found in other strains of
                canine adenovirus type 1; field isolate: Spibey et al.,
                Virus Research 14 (1989) 241-256."
                /citation=[3]
                /evidence=experimental
              join(493..1003,1101..1282)
                /function="transcription activation"
                /note="similar to human Ad E1a 298R protein; splice site
                by comparison with canine adenovirus type 2: Shibata et
                al., Virology 172 (1989) 460-467."
                /codon_start=1
                /product="E1a protein"
                /protein_id="AAB05429.1"
                /db_xref="GI:1477655"
                /translation="MKLLEAPRCLHEVVSQLEDQPCLSCEYSHGSGSPSPSLHD
                LFDLELRSPSPFLDCWCAADSESIETSDVGFNTLPVPSPSPSPSTSPASIP
                EDMLCLCEMTPTDGDVEVSATTSFEHWNEDFNPGVGFGLCKLCAIKYQKQGENSIC
                GYLKALAAEEVDADGEDDEIVFSAPKPGSKRSKSAVTSRGSVESSKARPCLEPEQT
                EPLDLSLRPPO"
              1406..1915
                /function="inhibition of apoptosis"
                /note="E1B 19kD protein"
                /codon_start=1
                /product="small T-antigen"
                /protein_id="AAB05430.1"
                /db_xref="GI:1477656"
                /translation="MDPLKICENYLTFRSIKGSTFSPGVFRWRFHADLVGVGNVIE
                REGRFWEIVPETHLTALWAFRGFTVAPTEILTSLQLENRGRQLAFLAFTSLLRNW
                PSDSVSEDAERLDLVCAWSRIQWSQAARLNDLPESVFGQGVVEERCGEHEHLA
                RSDDDPEFD"
              1768..3102
                /function="anti-oncogene"
                /note="E1B 55kD protein"
                /codon_start=1
                /product="large T-antigen"
                /protein_id="AAB05431.1"
                /db_xref="GI:1477657"
                /translation="MEQSDLESGRATNORPRVVRGAGVGRGVRERLASEGORR
                SPFLDQLQPLDSLXYTRAKLQRDHALEMPRGQVDFSLIEAREERAGPTDEWTFESVKT
                YRAKPGDGLATLKIKYAKLSEGVAEYKINSKIVTGACYYIIGNCALVRANLPVGTM
                FEVLNVDPISGIFMERIVFNSLIFDCRSTAVVCCISERNLTFHNCVFSGGHMLCLD
                IRAGAVRGCHFGVCAKLSKGLSVRVNRNSIFEKACAFGVSGSKASTSHSMFKDCA
                CCMFGSGQGTIANSHWMACTTTPMNLQCTCEGNSHVPLNGNTHFASNREAPMPT
                FNANLVRLYLRGGRGVPRKQSTFMCVIAAPRGVQVRIYLFVSVDYDCAIQLQG
                ENADVAERLCTGGRGINTSPSRAAYVTDTRIDREINSQDTAEFFSSDEDNL"
              3174..3485
                /function="virion stability"
                /codon_start=1
                /product="protein IX"
                /protein_id="AAB05432.1"
                /db_xref="GI:1477658"
                /translation="MDPQQKGIYNTCELTTRIPSWAGARONVTGSDLGGKVPSPDYLE
                SGRPLAAPRVNTLSEEQQLNNLITVNIIDDLKLTQVAAMQNSVTAIQEELKDLKQRYA
                R"
              complement(join(3502)..4829,5108..5120))
                /function="virion assembly"

```

AL
CE
RS
AL

S
ource

appea
appea

DS

DS

DS

DS

DS

AL
CE
RS
AL

S
ource

appea
appea

DS

DS

DS

DS

DS

AL
CE
RS
AL

S
ourc

appea
appea

DS

DS

DS

DS

DS

AL
CE
RS
AL

S
ourc

appea
appea

DS

DS

DS

DS

DS

CDS
RNERIERAVSGARQRELNDRMYRLREALTNSGLSEIEGGEDSARVRSKENWGAGAG
VGAASARYPHLLDYEEKENPDGVSFOQHERGTQSHENGHAEPAYSRQLGRFY"
10798..12489
/function="structural protein"
/note="Iiia protein precursor"
/codon_start=1
/product="piiaa"
/protein_id="AAB05437.1"
/db_xref="GI:1477663"
/translation="MRKRLTAPSHENSMSAALNPKMAAMQSQPTVDDSWAASISR
IMSTAGDRHKFSQPPANRLDALLEAVPSRADPTHEKVLITVINALIENGARRDEG
AGYDALLHRVYSINTOSNLNERLAEVRAQOVRIRAAAGNLSLTALNSFLARL
PAMVERQDNTGLFSLKLLVSEVNTVEIQSPHYFLQSSRNGTQTVNLNFAENL
KPLWGVKAPTEVETLSALTLNTRLLLVSPFTDSVLSRDSYGLVLLTYREALG
RNHLDMKLTVEVTEVSKMGNEINLQATLNFLLNRKKIKPKOYSLTPEERIVRF
IQOAVSLRMQENLSPTEALDVTANNEPFSVANNEDFINKLMDYFHRAALAPDIFFL
GAVNRPWLPPEGFFGVDPEDPDNYINDGDDSSLDLIRQDAMRFLKFKIDDDORT
ESRSVSRVPTPASSRRSSVAMASDSLIRPMNNDKSLREIEVLADKLARWTKYKRESE
EARSLEPVVVRPAKYSIAISSDESDDGMSKPKFLKFEKGGNPFALRPKLGRCL"
12340..13973
/function="structural protein; virion component III"
/codon_start=1
/product="penton base protein"
/protein_id="AAB05438.1"
/db_xref="GI:1477664"

CDS

Query Match 5.5%; Score 48.4; DB 16; Length 30288;
Best Local Similarity 50.4%; Pred. No. 0.0032;
Matches 118; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
QY 358 tgggctctccactgctattgtccagtcgccggtccctcagtgatagccgcggcgag 417
DB 28076 TGCAACTACATGTCACGTCTGACAGTTGCTTGGCAATGCTCTCTACGCAA 28017
QY 418 gtttggccagctggttaggtgggtggatggcccatgtttaatcagaggtttata 477
DB 28016 GTTTTAAAGAAATTTCTGAAGAATTTGTAATGGGTGGGTTGTGACAAAAGTATCTT 27957
QY 478 tggtagcgggagtggtgaattacaatgcaaaagaggttaattgttccagcgtg 537
DB 27956 TGGTACCGAGAGTTTGTGAATTTACAGACCTGTAGATTAACATGTTGGGGAGTATA 27897
QY 538 ttatgaggggtgcgcacttaactctacctgcgctgttggtatgatggccacgtg 591
DB 27896 ATGTTTAGAAACATTCATTACATTTACTTTAGACTGCTCTTTTTTAGCACCGTG 27843

Search completed: June 23, 2000, 09:34:53
Job time: 12475 sec

Measles virus 1983
Human secreted pro
HSV-2 strain SH5 C
O. sativa Xa21 gen
HSV-2 strain SH5 C
Sequence encoding
Cellular transcript
Measles virus Aik-
ceff (ORF9) encodi
Green fluorescent
Cepharmycin biosynt

35 32.2 3.6 15894 1 V18266
36 30.8 3.5 3301 1 V34209
37 30.8 3.5 6385 1 V62169
38 30.8 3.5 9424 1 X23525
39 30.8 3.5 117213 1 V62176
40 30.6 3.5 7326 1 Q99456
41 30.6 3.5 7326 1 V10093
42 30.6 3.5 15894 1 V18271
43 30.4 3.4 972 1 T00591
44 30.4 3.4 1908 1 V71079
45 30.4 3.4 2672 1 T08693

ALIGNMENTS

RESULT 1
T31315/c
ID T31315 standard; cDNA; 3189 BP.
AC T31315;
AT 16-APR-1997 (first entry)
DE Adenovirus E4 and part ITR region corresp. to bases 32749-35935.
KW Human adenovirus; inverted tandem repeat; E4; E1; Open reading frame;
KW expression; late gene; cell line; defective recombinant adenovirus;
KW complementation; gene therapy; cystic fibrosis; dystrophy; ds;
KW neurodegenerative disease; blood coagulation disorder; viral infection.
OS Adenovirus Ad5.
PN W09622378-A1.
PD 25-JUL-1996.
PF 19-JAN-1996; F00088.
PR 20-JAN-1995; FR-000747.
PR 01-JUN-1995; FR-006532.
PR 08-SEP-1995; FR-010541.
PA (RHON) RHONE POULENC RORER SA.
PI Dedieu J, Latta M, Orsini C, Perricaudet M, Vigne E;
PI Yeh P, Dedieu JF;
PI WPI: 96-354535/35.
DR Cells for prodn. of recombinant adeno and adeno-associated virus -
PT contain part of the E4 region of adeno virus, used to produce new
PT adeno virus defective in E1 and E4 regions for use as gene therapy
PT vectors
PS Disclosure: Page 37-39; 69pp; French.
PT This is the sequence corresponding to nucleotides 32749-35935 of the
CC human adenovirus type 5 (Ad5), which covers the E4 and right inverted
CC tandem repeat (ITR) region of the viral genome. The E4 region of the
CC genome contains 7 open reading frames (ORFs 1-7). These ORFs are thought
CC to regulate expression of the adenoviral late genes. The E4 region is
CC used to generate cell lines, esp. 293 cells, which can support the
CC production of defective recombinant adenoviruses, esp. those that lack
CC the E1 and part of the E4 region. The cell lines pref. have the E4
CC region inserted into their genomes for complementation of the defective
CC viruses. The E4 region is inserted under control of a promoter pref.
CC from MMV (Pharmacia) which is inducible by dexamethasone. Four defective
CC recombinant viruses are esp. mentioned in the specification: they all
CC lack all or part of the E1 region and part of the E4 region; virus AV1
CC also lacks nucleotides (nts) 34801-34329 and 34115-33126 consisting of
CC ORF3 and 6, AV2 also lacks all of the E4 region except for ORF4 (nts
CC 33093-33053), AV3 also lacks all of the E4 region except for ORF4
CC (present as nts 33093-33695 and 34634-35355) and AV4 further lacks the
CC ORFs of the E4 region i.e. nts 32720-35835, 33466-35355 or 33093-35355.
CC The defective viruses generated in the new cells can be used for gene
CC therapy e.g. for cystic fibrosis, dystrophy, neurodegenerative diseases,
CC blood coagulation disorders, viral infections, etc. 671 T;
SQ Sequence 3189 BP; 975 A; 912 C; 631 G; 571 T;

Query Match 100.0%; Score 885; DB 1; Length 3189;
Best Local Similarity 100.0%; Pred. No. 6.8e-269;
Matches 885; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 atgactacgtccgcggtccatttggcatgacactaacgacacagatcgtgtgtct 60
Db 1328 ATGACTACGTCCGCGGTCCATTGGCATGACACTAACGACACAGATCTCGTTGTCT 1269

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: June 23, 2000, 09:05:32 ; Search time 56.99 Seconds
(without alignments)
3885.243 Million cell updates/sec

Title: US-09-214-478-1
Perfect score: 885
Sequence: 1 atgactacgtccgcggtcc.....atgactacccccatgtag 885

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	885	100.0	3189	T31315	Adenovirus E4 and Plasmid pMTV6AORF.
2	885	100.0	3653	T59270	Second generation
3	885	100.0	8299	T59273	Chimeric plasmid p.
4	885	100.0	8299	V22144	Complete sequence
5	885	100.0	8710	V32370	Complete sequence
6	885	100.0	10610	V32375	Recombinant adenov
7	885	100.0	22026	T60559	Adenoviral vector
8	885	100.0	34303	V07261	Recombinant adenov
9	885	100.0	34382	X15627	Adenovirus vector
10	885	100.0	34427	X07371	Recombinant cis-ac
11	885	100.0	35000	T60557	Adenovirus 5 genom
12	885	100.0	35935	V07258	Recombinant trans
13	885	100.0	36538	T60558	Glucocorticoid rec
14	883.4	99.8	1884	T31320	Chimpanzee adenovi
15	875.4	98.9	36335	Q68003	Adenovirus 17. Rec
16	370	41.8	36519	V22141	Chimpanzee adenovi
17	319.2	36.1	35099	V27112	Complete DNA sequ
18	288.2	32.6	35524	V13168	Complete DNA sequ
19	49.8	5.6	6196	V13841	Complete DNA sequ
20	49.8	5.6	6243	V13169	Complete DNA sequ
21	49.8	5.6	6503	V13176	Complete DNA sequ
22	49.8	5.6	7379	V13176	Complete DNA sequ
23	49.8	5.6	8618	V18741	Complete DNA sequ
24	49.8	5.6	8792	V18745	Complete DNA sequ
25	37.8	4.3	34446	X24332	Bovine adenovirus
26	33.8	3.8	15894	V18267	Measles virus Mont
27	33.4	3.8	502	Q65083	Human Id-1H' gene.
28	33.4	3.8	509	Q65082	Human Id-1H gene.
29	32.2	3.6	1391	V72242	Human DPL homologu
30	32.2	3.6	15894	V18264	Measles virus Edmo
31	32.2	3.6	15894	V18268	Measles virus Rube
32	32.2	3.6	15894	V18269	Measles virus Mora
33	32.2	3.6	15894	V18270	Measles virus Zagr
34	32.2	3.6	15894	V18265	Measles virus 1977

	Pd	12-DEC-1996.	U10245.	
Pf	04-JUN-1996;	US-462014.		
Pr	05-JUN-1995;	US-462014.		
Pa	(UYPE-) UNIV PENNSYLVANIA.			
Fisher KJ,	Gao G,	Willson JW;		
RPI:	97-043152/04.			
DRecombinant adeno-associated virus contg. second gene which facilitates its conversion from single stranded to double stranded virus - enhances efficiency of ex vivo transduction into target cell Example 19; Page 108-114; 13ipp; English. This is the nucleotide sequence of a second generation of recombinant adeno-associated virus (rAAV) which contains 2 genes, the first gene able to express the gene of interest (beta-galactosidase) in the target cell; and the second gene(Ad5 E4 ORF6) able to convert the single stranded virus to its double stranded form upon expression. The rAAV is useful in pharmaceutical compositions for use in ex vivo and in vivo genetic therapy for the treatment of inherited diseases, cancer and other genetic dysfunction, e.g. cystic fibrosis. See also T59271 for another example of a recombinant adeno-associated virus.				
SQ	Sequence	8299 BP;	1816 A;	2130 C;
			2108 G;	1963 T;
	Query Match	100.0%; Score 885;	DB 1;	Length 8299;
	Best Local Similarity	100.0%;	Pred. No. 1.le=268;	Mismatches 0;
	Matches	885;	Conservative	Indels 0;
			Gaps	0;
Qy	1	atgaacagctcggcggttcatttgcgatgacctcacaccacacacacagtccgttgtct	60	
Db	3402	ATGACTACGTCCGGCGTTCATTGGCATGCATCAGCACCAACACGATCTCGTGTTCT	3461	
Qy	61	cggcgcaactccgttacagttagtgaggatcctaactccccttttgagaagaaacccgccctacc	120	
Db	3462	CGGCGCACTCCGTACAGTAGGATCGTCACTTCCTTTTTGACACAGAACAACC GGCGCTACC	3521	
Qy	121	atactgagcatcatccgcgtgcgcgaagttaaacatttgacaatgcacaacgtgagt	180	
Db	3522	ATATGAGGANTATCGCTGCTGCCGAATGTAAACATTTGACATAATGACACACGTGAGT	3581	
Qy	181	tacgtcgaggtctccccctgcagtggtggttatctacgtgattcaggaatgggttgctcc	240	
Db	3582	TACGTGGAGGTCTTCCTCAGTGTGGGATTTACGCTGATTGAGGAATGGGTGTGTCCC	3641	
Qy	241	tggatatgtttaacgcggggaggccttgtaaactccctgaggagtgatgcacgtgtgc	300	
Db	3642	TGGGATATGGTTCTAACCGGGAGGAGCTTGTAACTCTGAGGAAGTGTATGCACGTGTGC	3701	
Qy	301	ctgtgtgtgccaatgtatcatcatgacagagatgatgatcatcattggtttacagtccttg	360	
Db	3702	CTGTGTTGTCCAACATTGATATCATGACGAGCATGATGATCCTGTTTACGAGTCTCTGG	3761	
Qy	361	gctcceaactcattgtccagtcgccggttcctcagtgtagttagccggcgggcaggtt	420	
Db	3762	GCTCTCCACTGTCAITGTTCAGTCCCGGTCCCTGCAGTGTATGATAGCGCGGCAGGTT	3821	
Qy	421	ttggccagctggttttaggatggtgtggtggatggcgccaattggttaatacagaggtttat	480	
Db	3822	TTGGCCAGCTGGTTTAGGATGGTGTGGATGGCGCCATGTTTAATCATGAGAGGTTTATAGG	3881	
Qy	481	taccgggagtggtgaattaaccaatgcacaaagaggttaattgtttatgtccagcgtgttt	540	
Db	3882	TACCGGGAGGTGGTGAATTACAAATGCCAANAAGAGGTAAATGTTTATGTCCAGCGTGT	3941	
Qy	541	atgaggggtgcgcaacttaactacccctgcgtgtggtatgtatgccccagtggtttctgtg	600	
Db	3942	ATGAGGGGTGCCACTTAATCTACCTGCGCTTGTGTTATGATGCCACGTGGGTCTCTGTG	4001	
Qy	601	gtccccgcatagcttttgatacacagcgccctgcactgtggagatttgaaaacaattgtg	660	
Db	4002	GTCCCCGCCATGAGCTTTGGATACAGCGCTTTGCACCTGTGGGATTTTGNACAAATATCTG	4061	
Qy	661	gtgctgtgctgagttactgtgctgatttaagttagatcagggtgcgtgctgtgccccg	720	
PN	W09639530-A2.			

b	30168	ATGACTACGTCCGGCGTTCCATTGGCATCAGCACTACGACCAACGATCTCGGTGTCT	301109
y	61	cggcgcaactccggtacagtaggatcgtctacctctttagagacgaaccgcgctacc	120
b	30108	CGGCGCACTCCGTACAGTAGGGAATCGTCTACCCTCTTTTGAGACAAGAACCCGCCCTACC	30049
y	121	atactgaggatcacccgctcgtcccgaatgtaaacactttgacaattcacaaactgagt	180
b	30048	ATACTGAGAGATCATCCCCTGCTCCCGCAATGTAACACTTGAATGCACAACTGAGT	29989
y	181	tacgtcgagggtcttcctcgtcagtgtaggatttacgctgattcaggaatgggttatcccc	240
b	29988	TACGTCCGAGGTCTTCCTCGAGTGTGGATTTACGCTGATTCAGGAATGGGTGTGTCCC	29929
y	241	tgggatatggttctaaccgagagagccttgaatcctgagggaagtgtatgcacgtgtgc	300
b	29928	TGGGATATGGTCTTACGCGGAGAGAGTTGTAACTCCTGAGGAAGTGTATGCACGTGTGC	29869
y	301	ctgtgttgtccaacaattgatcatgacgagcagatgatcatcoattggttacgagctcgg	360
b	29868	CTGTGTGTGCGCAACATTCATATCATGACGASCATGATGATCCATGGTTACGAGTCTCTGG	29809
y	361	gctctcaactgcatgttccagtcgccggttccctgcagtgtagcggcggcgcsaggtt	420
b	29808	GCTCTCACATGTCATTGTTCACGTCCCGGTTCCTCTGCAGTGTATAGCGCGGCGCAGGTT	29749
y	421	ttagccagctgggttttaggattggttggatggcgccatggtttaatcagagatttatgg	480
b	29748	TTGCCAGCTGGTTTTAGGATGTTGTGTGATGCGGCCAATGTTTAATCAGAGTTTATATGG	29689
y	481	taccgggaggtggtgaattacaatgcaccaagaggaatgatttatctgcacgcgtgttt	540
b	29688	TACC GGAGGTGGTGAATTAACAATGCCAAGAAGGTAATGTTTATGTCACAGCTGTTT	29629
y	541	atgaggvggtcccaacttaactaccctgcgcttggtgatgagccagtggttctcgtg	600
b	29628	ATGAGGGTCCCACTTAATCTACCTTCGCTTGTGATGATGAGGCCAGTGGGTCTCTGTG	29569
y	601	gtcccgcacatgaacttgatcacagcgccttcgacttgggattttgaacaaatattgtg	660
b	29568	GTCCCCGCCATGAGCTTTGGATACAGCCCTTCGACTGTGGATTTGAACAATATTGTG	29509
y	661	gtgctgtgctgcagttactgtgtgattaaagttagatcagggctgctgtgtgcocgg	720
b	29508	GTGCTGTGCTGCAGTTACTGTGCTGATTAAAGTGAGATCAGGGTGGCTGTGTGCCCGG	29449
y	721	aggacaaggcgcttatctcggcgcggtgcgaatcatcgctgagagaccactgccatg	780
b	29448	AGSACAAGCGCCTTATGCTCGCGCGGTGGCAATCATCGCTGAGGAGACCACCTGCCATG	29389
y	781	ttgtattcctgcaggacgagcggcgcgcgagcagttattcgcgcgtgctgtgagcac	840
b	29388	TTGTATTCTGCAGGACGGAGCGCGCGGCGCAGCAGTTATTGCGCGCTGTGTGAGCAC	29329
y	841	caccgccttatctgatgcagattatgactctacccccatgtag	885
b	29328	CACCGCCCTATCTGTATGCAGCATTTATGACTCTACCCCCATGTAG	29284
RESULT 8			
V07261/c			
ID	V07261 standard; DNA; 34303 BP.		
DT	28-SEP-1998 (first entry)		
DE	Adenoviral vector plasmid pBHgl1.		
KW	Adenovirus 5; Ad5; vector; gene therapy; ds.		
OS	Mastadenovirus 5.		
FH	Key Location/Qualifiers		
FT	misc_feature /tag= "a"		
FT	/note= "region deleted in plasmid pdePTP and pBHgl1delPRP (Claim 45)."		
FT	misc_feature 8773..9586		

FT		b	/*tag=	"region deleted in plasmid	
FT		/note=	pAXBdelpoldeletPPVARNA+tl3 and		
FT			pBHGlldelpolydeletPPVARNA+tl3	(Claim 53)"	
FT			.11314		
FT		C	10705.		
FT		/*tag=	"region deleted in plasmid pdeTP and		
FT		/note=	pBHGlldelptP (Claim 51)"		
FT			.12513		
FT		D	11067.		
FT		/*tag=	"region deleted in plasmid		
FT		/note=	pAXBdelpoldeletPPVARNA+tl3 and		
FT			pBHGlldelpolydeletPPVARNA+tl3	(Claim 53)"	
FT					
PN			WO9817783-A1.		
PD			30-APR-1998.		
PF			23-OCT-1997; UI9541.		
PR			23-OCT-1996; US-735609.		
PA			(UNWI) UNIV MICHIGAN		
PI			Amalfitano A, Chamberlain JS, Hartigan-O'Connor DJ,		
PI			Hausner MA, Kumar-Singhnr;		
DR			WP1; 98-261485/23.		
PT			New adenoviral recombinant plasmid(s) - comprise sequences provided		
PT			for expression of large foreign DNA fragments, used for, e.g. gene		
PT			therapy of genetic disease(s)		
PS			Claim 45; Page 87-102; 13ppp; English.		
CC			This nucleotide sequence comprises plasmid pBHGl1 that consists of		
CC			An EI-deleted adenovirus (Ad) genome. It contains a deletion of		
CC			Ad5 (see also V07258) from bp 188-1339 (0.5-3 m.u.). This		
CC			deletion removes the packaging signals as well as EI sequences.		
CC			pBHGl1 also contains a large deletion within the E3 region (bp		
CC			27865-30995, 77.5-86.2 m.u.). Nucleotide 8773 of pBHGl1 is		
CC			equivalent to nucleotide 7269 of Ad5. The large E3 deletion		
CC			provides a large cloning capacity to Ad vectors based on pBHGl1.		
CC			Recombinant plasmids pdeIpol and pXBdelpoldeletPPVARNA+tl3 and		
CC			polymerase functions) and pAXBdelpoldeletPPVARNA+tl3 and		
CC			pBHGlldelpoldeletPPVARNA+tl3pBHGl1 (deleted for EI, polymerase as		
CC			preterminal protein functions), are specifically claimed. The		
CC			invention provides improved adenoviral vectors and packaging		
CC			cell lines. One type of improved vector comprises deletions		
CC			within the E2b region of the adenoviral genome (see also V07261).		
CC			These E2b-deleted virus are used in conjunction with novel cell		
CC			lines that constitutively express E2b gene products. The invention		
CC			also provides vectors deleted for all viral coding regions. These		
CC			'guttred' vectors permit the transfer of large genes (e.g. up to		
CC			35 kb) to cells, as demonstrated by the transfer of the dystrophin		
CC			gene to the muscle of mice. The E2b-deleted and gutted vectors		
CC			provide improved adenoviral vectors useful for a wide variety of		
CC			gene therapy applications.		
CC			Sequence 34303 BP; 7948 A; 9625 C; 9517 G; 7213 T;		
SQ					
					Query Match 100.0%; Score 885; DB 1; Length 34303;
					Best Local Similarity 100.0%; Pred. No. 2,1e-268;
					Matches 885; Conservative 0; Mismatches 0; Indels 0; Gaps
Qy	1	atgaactacgtccgcgcttccattggcgatgacatacagaccacaacagatcgtgttcct	60		
Db	32446	ATGACTACGTCCGGCGTTCCATTGGCATGCACCTACGACCACACATCTCGGTTGTC	32387		
Qy	61	cggcgactccgtacagtaggatcgtctacctcttttgacagagaaccccgctacc	120		
Db	32386	CGGCGCACTCCGTACAGTAGGATGCTTACCTCCTTTTGACAGAACAACCGCGCTACC	32327		
Qy	121	atactggagatcatccgctgctgcccgaaatgaacactttgacaatgcacacgtag	180		
Db	32326	ATACTGGAGGATCATCCGCTGCTGCCGAATGTAACTTTGACAATGCACAACGTGAT	32367		
Qy	181	tacgtcgaggtcttccctcagtgaggattacgctgattcaggaaatggttgtcccc	240		
Db	32266	TACGTGCGAGGTCTTCCCTCAGTGTGGGATTTACGCTGATTCAGGAATGGTTGTTCCC	32207		
Qy	241	tgggatattggttctaacgcggggaggagctgttaactcctgaggaatgatacagtg	300		

Db 32206 TGGGATATGTTCTTACGCGGAGGAGCTTGAATCCTGAGGAAGTATGACAGCTGTC 32147
 QY 301 ctgtgtgtgccaacattgatcatgacgagcatgatgatccatggttacgaatctgtg 360
 Db 32146 CTGTGTGTGCGCAACATTGATATCATGACGACATGATGATCCATGTTACGATCTCTGG 32087
 QY 361 gctctccactgtccattgttccagttcccggttccctgcagtgatagccggcgaggtt 420
 Db 32086 GCTCTCCATGTCATGTTTCCAGTCCCGGTTCCCTGCACTGATATAGCGGGGACAGTT 32027
 QY 421 ttggccagtggttttagtagtggtgagtgagcgcccatggttttaacagaggtttatgg 480
 Db 32026 TTGGCCAGCTGTTTATGAGATGCTGTGATGCGGCCCATGTTTATACAGAGTTTATATG 31967
 QY 481 taccggaggtgtgaattacaacatgcacaaagaggttaattgtttatgtccagcggttt 540
 Db 31966 TACCGGGAGGTGTGATTAACAATGATCCAAAGAGGTAATGTTATGTCCAGCGTGT 31907
 QY 541 atgaggggtgcgaccttaactactgcgtgctgtgtgtgtatgacgccaagtgttctgtg 600
 Db 31906 ATGAGGGGTGCGCACTTAATCTACTGCGCTTGGTGTATGATGCCACGTGGTCTGTG 31847
 QY 601 gtcccccagtgagctttgtatagacagcgcttgctgctggtggttttgaacaattgtg 660
 Db 31846 GTCCCGCCATGAGCTTGGATACAGCGCTTGCACTGTGGGATTTGAACATATGTG 31787
 QY 661 gtgtgtgtgtgagcttactgtgctgattgaatgaatgagatcaggtgctgtgtgcccgg 720
 Db 31786 GTGTGTGTGCTGAGTACTGTGCTGATTAAAGTGAGATCAGGTGCGCTGTGCCCCG 31727
 QY 721 agcaagagggccttactgtgcggggtgcggaatcatcgtgagagaccactgccatg 780
 Db 31726 AGCAAGGCGCTTATGCTGCGGGGGTGGCGAATCATCGCTGAGGAGACCACTGCCATG 31667
 QY 781 ttgtattcttcgagcgagcgagcgcgccgagcagctttatgcgcgtgctgcagcac 840
 Db 31666 TTGTATTCTGACGAGCGAGCGCGCGCGGCGACAGTTTATTTCGCGCGCTGCTGAGCAC 31607
 QY 841 caccgacctatcgtatgcacgattatgactactacccccatgtag 885
 Db 31606 CACCGCCCTATCCTGATGACGATTATGACTCTACCCCATGTAG 31562

RESULT 9

X15627/c
 ID X15627 standard; DNA; 34382 BP.
 AC X15627;
 DT 07-MAY-1999 (first entry)
 DE Recombinant adenovirus Ad:Pac-beta-Gal.
 KW Recombinant adenovirus Ad:Pac-beta-Gal; replication deficient adenovirus;
 KW Ela region; E3 region; gene therapy; nitric oxide synthetase; NOS;
 KW cystic fibrosis chloride channel; LDL receptor; erythropoietin;
 KW atherosclerotic artery; ss.
 OS Synthetic.
 OS Mastadenovirus.
 FH Key
 FT CDS
 FT complement (3372..333)
 FT /*tag- a
 PN US5880102-A.
 PD 09-MAR-1999.
 PF 17-JAN-1995; 374483.
 PR 17-JAN-1995; US-374483.
 PA (UYDU-) UNIV DUKE.
 PI Blazing MA, George SE;
 DR WPI; 99-204005/17.
 DR P-PSDB; W97243.
 PT New replication deficient adenovirus bearing deletions of the Ela
 PT and E3 regions - containing a single packaging signal sequence and
 PT Ela enhancer sequence, the Ela deletion has unique cleavage sites
 PT and is useful as a gene therapy vector.
 PS Disclosure; Fig 49-100; 148pp; English.
 CC The present sequence represents recombinant adenovirus
 CC Ad:Pac-beta-Gal, which exemplifies the invention. The specification

CC describes replication deficient adenoviruses having deletions of
 CC the Ela and E3 regions and comprise a single packaging signal sequence
 CC and Ela enhancer sequence, where the sequences are at the 3' end of the
 CC adenovirus and the Ela deletion contains at least one PacI, ClaI, XbaI
 CC or BstBI cleavage site. The replication deficient viral vectors can be
 CC used in gene therapy regimens to effect the transfer of genes encoding
 CC molecules of therapeutic importance, including isoforms of the nitric
 CC oxide synthetase (NOS) gene (brain, endothelial and microphage NOS), the
 CC cystic fibrosis chloride channel (CFTR) gene, the dystrophin gene, the
 CC LDL receptor gene and the erythropoietin gene. The NOS isoforms can be
 CC used in vascular applications or in cancer therapy (microphage NOS). The
 CC NOS gene can be introduced into vein grafts prior to their use as
 CC coronary artery bypass grafts. A NOS containing adenovirus can also be
 CC used following coronary angioplasty to prevent restenosis and to treat
 CC atherosclerotic arteries.
 SQ Sequence 34382 BP; 7923 A; 9880 C; 9421 G; 7158 T;

Query Match 100.0%; Score 885; DB 1; Length 34382;
 Best Local Similarity 100.0%; Pred. No. 2.1e-268;
 Matches 885; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgactacgtccgcggttccatttggcatgacactacgacacacagatctcggtgtgt 60
 Db 32278 ATGACTACGTCCGCGGTCCATTGGCATGACACTAGACCAACACGATCTCGGTGTGT 32219
 QY 61 cgcgcgactccgtacagtagggatcgtctacctctctttgagacagaaacccgcgtacc 120
 Db 32218 CGCGCACTCCGTACAGTAGGGATCTCTACCTCTTTTGAGACAGAAACCCGCGTACC 32159
 QY 121 atactgagatcatcgcgtgctgcccgaatgtaacatttgaacatgacgaactgagt 180
 Db 32158 ATACTGAGGATCATCGCTGCTGCCGAATGTAACTTTTACCAATGCATAACGTGAGT 32099
 QY 181 taagtgcgaggttccctgcagtggtgattacgctgattcaggaatggtgtgtccc 240
 Db 32098 TAGCTGCGAGGTCTTCCCTGCACTGTGGATTTACGCTGATTACAGAAATGGTGTTCCT 32039
 QY 241 tgggatatggttctaaccgagcgagcttgaatcctgaggaagtgatgcaggtgtgc 300
 Db 32038 TGGGATATGTTCTAACCGGGGAGGAGCTTTGATCTCCGAGAAAGTGTATCAGGTGTGC 31979
 QY 301 ctgtgtgtgccaacattgatatacagcagcatgatgacctggtttacgaatcctgg 360
 Db 31978 CTGTGTGTGCGCAACATTGATATCATGACGACATGATGATCCATGTTACGATCTCTGG 31919
 QY 361 gctctccactgtcattgtccagtcggttccctgcagtgatagccggcgaggtt 420
 Db 31918 GCTCTCCACTGTCAATGTTCAGTCCCGGTTCCCTGAGTGTATAGCGCGGACAGTT 31859
 QY 421 ttggccagctggttttaggagtggtggtgagtgcccatggttttaacagaggtttatgg 480
 Db 31858 TTGGCCAGCTGTTTAGGATGCTGTGTGATGATGCCACGTGGTGTCTGTG 31799
 QY 481 taccggaggtgtgaattacaacatgcacaaagaggttaattgtttatgtccagcggttt 540
 Db 31798 TACCGGGAGGTGTGATTAACAATGATCCAAAGAGGTAATGTTATGTCCAGCGTGT 31739
 QY 541 atgaggggtgcgaccttaactactacgtgcgtctgtgtatgacgccaagtgttctgtg 600
 Db 31738 ATGAGGGGTGCGCACTTAATCTACCTGCGCTTGGTGTATGATGCCACGTGGTCTGTG 31679
 QY 601 gtcccccagtgagctttgtgatacagcgcttgcactgtgggattttgaacaattgtg 660
 Db 31678 GTCCCGCCATGAGCTTGGATACAGCGCTTGCACTGTGGGATTTGAACATATGTG 31619
 QY 661 gtgtgtgtgcagttactgtgtgtattgaatgagatcaggtgctgtgtgtgcccgg 720
 Db 31618 GTGTGTGTGCTGAGTACTGTGCTGATTAAAGTGAGATCAGGGTGTGCTGTGTGCCGG 31559
 QY 721 agcaagagggccttactgtgcggggtgcggaatcatcgtgagagaccactgcctg 780
 Db 31558 AGCAAGGCGCTTATGCTGCGGGGTGCGAATCATCGCTGAGGAGACCACTGCCATG 31499

KW	Adenovirus H5. 020CMVgag-pol; plasmid pAdMLVneo-int; Ad5; Mo-MLV;
KW	retrovirus; retrotransposition; transposase; transgene;
KW	gene therapy; vector; ss.
OS	Chimeric - Mastadenovirus serotype 5.
OS	Chimeric - Moloney murine leukaemia virus.
OS	Chimeric - Cytomegalovirus.
PN	WO9715679-A1.
PD	01-MAY-1997.
PD	01-MAY-1997.
PD	24-OCT-1996; US-005942.
PR	27-OCT-1995; US-005942.
PA	(UYPE-) UNIV PENNSYLVANIA.
PI	Kelley WM, Wilson JM;
PT	WPI: 97-259031/23.
PT	Recombinant replication defective virus - inserts transgene into
PT	host cell chromatin in the presence of transposase, providing stable
PT	and durable transgene expression
PS	Example 2; Fig 9A-I; 7app; English.
CC	This is the DNA sequence of the trans-acting recombinant adenovirus
CC	H5.020CMVgag-pol. It was obtained by co-transfecting HEK293
CC	(ATCC CRL1573) cells with linearised plasmid pAdCMVgag-pol (see
CC	T6055) and Ad5dl7001. A cis-acting adenovirus (see T60557) was
CC	also produced, and retrotransposition and integration into host
CC	cells was demonstrated in HeLa cells infected with both viruses.
CC	Claimed novel recombinant replication defective viruses (A)
CC	comprise: (a) DNA of, or corresponding to, at least part of the
CC	viral genome, able to infect a mammalian cell; and (b) a first
CC	expression sequence (ESI) comprising a human gene (I) linked to
CC	regulatory sequences for expression, with the cassette flanked by
CC	the cis-acting terminal repeat sequences of a transposon, and ESI
CC	flanked by DNA of (a); (A) can infect a mammalian cell and can
CC	express (I) and transfer it to the cellular chromatin in vivo or in
CC	vitro in the presence of a transposase. Also new are: (1) a
CC	recombinant replication defective virus (B) containing (A) as above
CC	and a second expression sequence (ES2) containing a trans-acting
CC	transposase gene (II) plus regulatory sequences, flanked by DNA of
CC	(a) and able to express transposase in mammalian cells; and (2)
CC	mammalian cells which stably express (I) integrated into its
CC	chromatin, produced by infection with (A). (A) and (B) are used
CC	for delivery and stable integration of (I) into a host cell
CC	(claimed), e.g. in somatic gene therapy of genetic defects or
CC	deficiencies such as cystic fibrosis. (A) are also used to prepare
CC	recombinant retroviruses (RVV) (claimed). Because (I) becomes
CC	stably integrated, it provides longer lasting expression than genes
CC	introduced with conventional adenoviral vectors and the need for
CC	repeated administration is avoided. The transgene is inherited by
CC	progeny cells and the viral component is gradually degraded by the
CC	cell. When used for RVV production, (A) increase the ratio of RVV
CC	to empty retrovirus.
CC	Sequence 36538 BP; 8621 A; 10490 C; 9909 G; 7518 T;
QY	Query Match 100.0%; Score 885; DB 1; Length 36538;
DB	Best Local Similarity 100.0%; Pred. No. 2.2e-268; Indels 0; Gaps 0;
DB	Matches 885; Conservative 0; Mismatches 0;
QY	1 atgactacgtccggttcattggcattgacacactacgacacacacacgtctcgttctct 60
DB	34680 ATGACTACGTCCGCGGTTCATTGGCATGACACTAGACACACACGATCTCGTTGTCT 34621
QY	61 cggcgactccgtacagtaggagtcgtctacctctcttttgagacagaaacccgcgtacc 120
DB	34620 CGGCGCACTCCGTACAGTAGGGATCGTCTACTCTCTTTTGAGACAGAAACCCGCGTACC 34561
QY	121 atactgaggatcatccgctgctccggaatgtaacactttgacaaatgcacaacgtgagt 180
DB	34560 ATACTGGAGGATCATCCGCTGCTGCCGAATGTAACTTTGACAAATGACAACTGAGT 34501
QY	181 tacgtgcaggtcttccctgcagtggtgagatttacgtgatttcaggaatgggtttctccc 240
DB	34500 TACGTGCGAGGCTCTCCCTGCGAGTGGGATTTACGCTGATTAGGAATGGTGTGTTCC 34441
QY	241 tgggatattggttctaaccgaggagagcttctaactcctcaggaagtgtatcaagtgtagc 300
QY	100.0%; Score 885; DB 1; Length 35935;
QY	Best Local Similarity 100.0%; Pred. No. 2.1e-268;
QY	Matches 885; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 atgactacgtccggttcattggcattgacacactacgacacacacacgtctcgttctct 60
DB	34077 ATGACTACGTCCGCGGTTCATTGGCATGACACTACGACCAACAGATCTCGGTTGTCT 34018
QY	61 cggcgactccgtacagtaggagtcgtctacctctcttttgagacagaaacccgcgtacc 120
DB	34017 CGGCGCACTCCGTACAGTAGGGATCGTCTACCTCTTTTGAGACAGAAACCCGCGTACC 33958
QY	121 atactgaggatcatccgctgctgcgcgaatgtaacactttgacaatgcacaacgtgagt 180
DB	33957 ATACTGGAGGATCATCCGCTGCTGCCGAATGTAACTTTGACAAATGCAACAGTGAGT 33898
QY	181 tacgtgcaggtcttccctgcagtggtggtattacgctgattcaggaatgggttctccc 240
DB	33897 TACGTGCGAGGCTCTCCCTGCGAGTGGGATTTACGCTGATTACGGAATGGTGTGTTCC 33838
QY	241 tgggatattggtctaaacgaggagagcttctaactcctcgtgaggaagtgtatgcacgtgac 300
DB	33837 TGGGATATTGTTTACACGGCGGAGAGCTTGTAATCTGAGGAAGTGTATGCAGCTGTC 33778
QY	301 ctgtgtgtgccaacattgatcatcagcagcagatgatccatggttacagagctctgg 360
DB	33777 CTGTGTTTGCCAAACATTGATATCATGACGACGATGATGATCCATGTTTACGAGTCTCGG 33718
QY	361 gctctccactctattgtccagtcctcgttccctcgtgattagcggcgaggcaggtt 420
DB	33717 GCTCTCCACTGTCATGTCTCCAGTCCCGGTTCCCTGCGAGTATAGCCGCGGCGCAGGTT 33658
QY	421 ttggcgagctggtttagtagtggtggtgagcgcacattgtaactcgtgaggaagtgtatgcacgtgac 480
DB	33657 TTGGCCAGCTGTTTAGGATGTTGTTGATGGCGCATGTTTAACTACAGAGTTTATATGG 33598
QY	481 taccggggaggtggtgaattacaacatgcacaaagaggaatgtttatctcagcgtgttt 540
DB	33597 TACCGGGAGGTGGTGAATTACAAACATGCCAAAGAGGTAATGTTTATGTCAGCGGTGTT 33538
QY	541 atgaggggtcgcacattacatctacgtcgtctgtggtatgagccacgtgggttctctg 600
DB	33537 ATGAGGGGTGCGCACTTAATCTACCTGCGCTTGTGTATGATGTCGACAGTGGGTCTGTG 33478
QY	601 gtcccgccatgagctttggatcacagcgtctgacgtggtgattttgaacaattctg 660
DB	33477 GTCCCGCCATGAGCTTTTGGATACAGCGCTTGCATGTTGGGATTTTGAACAAATATTGTG 33418
QY	661 gtactgtgctgactgactgctgatttaagtgagatcaggtgctgctggtgccccgg 720
DB	33417 GTGCTGTGTGAGTACTGTCTGATTTAAGTAGATCAGGCTGCTGTGTCGCGG 33358
QY	721 aggaacagggcgccttatgtcgtggggtggtgcgaatcgtcgtgagagacacactccatg 780
DB	33357 AGGACAAGCGCCTTATGCTGCGGCGGTGCGAATCATCGCTGAGGAGACCACTGCCATG 33298
QY	781 ttgtattctcagagcagagcggcgccgagcagcaggtttatttcgcgcgtctgcagcac 840
DB	33297 TTGTATTCTTGCAGGACGAGCGGCGGCGGACAGTATTATTCGCGCGGTGTGTGACGAC 33238
QY	841 caccgcctatctgacgcagattatgactctaccccccatgtag 885
DB	33237 CACGCGCTATCTGATGACAGTATTATGACTCTACCCCATGTAG 33193
RESULT	13
T60558/c	
ID	T60558 standard; DNA; 36538 BP.
AC	T60558; 1998 (first entry)
DT	05-MAR-1998
DE	Recombinant trans-acting adenovirus H5.020CMVgag-pol.

Mon Jun 26 09:10:29 2000

```
Qy 661 gtgctgtgtcgcagttactgtgatttaagttagatcagggcgctgtgctgcccgg 720
Db 1381 gfgctgtgtcgcagttactgtgctgatttaagttagatcagggcgctgtgctgcccgg 1440
Qy 721 aggacaagcgcccttatctgtgcccgtgtggaatcatcgtgagagaccactgcccag 780
Db 1441 aggacaagcgcccttatctgtgcccgtgtggaatcatcgtgagagaccactgcccag 1500
Qy 781 ttgtattctcgcagacgagcgcgcgcgagcagcagcttatttcgcgcgctgtgagcac 840
Db 1501 ttgtattctcgcagacgagcgcgcgcgagcagcagcttatttcgcgcgctgtgagcac 1560
Qy 841 caccgcccctctctgacgacgattatgacttaccctaccctaccctag 885
Db 1561 caccgcccctctctgacgacgattatgacttaccctaccctaccctag 1605

RESULT 15
Q68003/c
ID Q68003 standard; DNA; 36335 BP.
AC Q68003;
DT 26-MAR-1996 (first entry)
DE Ad2/ORF6/PGK-CFTR nucleotide sequence.
KW Recombinant adenovirus; Ad2/CFR6-1; adenovirus 2 serotype; Ela; Elb;
KW viral replication; gene expression; gene therapy; cystic fibrosis;
KW cystic fibrosis transmembrane conductance regulator; CFTR;
KW promoter; E3; p19; MHC; class 1; viral latency; pulmonary airway; ds.
OS Synthetic.
FH Key
FT misc_feature
FT 12915..36335
FT /tag= a
FT /note= "Represents residues 10676-34096 of Ad2-E4/ORF6"
FT misc_feature
FT 35069..35973
FT /tag= b
FT /note= "Represents residues 33178-34082 of Ad2"
FT misc_feature
FT 12915..35054
FT /tag= c
FT /note= "Represents residues 1-32815 of Ad2"
FT exon
FT 28478..28790
FT /tag= d
FT /number= 1
FT /note= "33K protein"
FT intron
FT 28791..28992
FT /tag= e
FT /number= 1
FT /note= "33K protein"
FT exon
FT 28993..29366
FT /tag= f
FT /number= 2
FT /note= "33K protein"
FT cds
FT 13279..14526
FT /tag= g
FT /product= 52.55K protein
FT cds
FT 14547..16304
FT /tag= h
FT /product= IIXa protein
FT /note= "peripentonal hexon-associated protein"
FT polya_signal
FT 16331..16336
FT /tag= i
FT /note= "Major late mRNA L1 poly A signal (putative)"
FT cds
FT 16390..18105
FT /tag= j
FT /product= Penton protein
FT /note= "viral component III"
FT cds
FT 18112..18708
FT /tag= k
FT /product= Pro-VII protein
FT /note= "precursor to major core protein"
FT cds
FT 18778..19887
FT /tag= l
FT /product= pV protein
FT /note= "minor core protein"
FT polya_signal
FT 20186..20193
FT /tag= m
```

```
FT cds
FT 20240..20992
FT /tag= n
FT /product= pVI protein
FT /note= "Hexon associated precursor"
FT cds
FT 21077..23983
FT /tag= o
FT /product= Hexon protein
FT /note= "viral component II"
FT polya_signal
FT 24657..24662
FT /tag= p
FT /note= "Major late mRNA L3 poly-A signal (putative)"
FT cds
FT 24729..26318
FT /tag= q
FT /product= DBP protein
FT /note= "DNA binding or 72K protein"
FT cds
FT 26347..28764
FT /tag= r
FT /product= 100K protein
FT /note= "Hexon assembled"
FT cds
FT 29454..30137
FT /tag= s
FT /product= pVIII protein
FT /note= "Hexon-associated precursor"
FT polya_signal
FT 30444..30449
FT /tag= t
FT /note= "Major late mRNA L4 poly-A signal (putative)"
FT cds
FT 31051..31530
FT /tag= u
FT /product= E3 19K protein
FT /note= "Glycosylated membrane protein"
FT cds
FT 31707..32012
FT /tag= v
FT /product= E3 11.6K protein
FT polya_signal
FT 32008..32013
FT /tag= w
FT /note= "E3-1 mRNA poly-A signal (putative)"
FT polya_signal
FT 33081..33086
FT /tag= x
FT /note= "E3-2 mRNA poly-A signal (putative)"
FT polya_signal
FT 35013..35018
FT /tag= y
FT /note= "Major late mRNA L5 poly-A signal (putative)"
FT misc_feature
FT 1..12914
FT /tag= z
FT /note= "Represents residues 1-12914 of pAd2/PGK-CFTR"
FT promoter
FT 380..914
FT /tag= aa
FT /note= "pgk promoter"
FT cds
FT 1011..5453
FT /tag= ab
FT /product= CFTR
FT W09412649-A.
FT 09-JUN-1994.
FT 02-DEC-1993: U11667.
FT 03-DEC-1993: US-985478.
FT 01-OCT-1993: US-130682.
FT 13-OCT-1993: US-136742.
FT (GENZ ) GENZYME CORP.
FT PA Armentano D, Couture LA, Gregory RJ, Smith AE;
FT WPI: 94-200277/24.
FT Adeno:virus-based gene therapy vectors - esp. useful for gene
FT therapy of cystic fibrosis
FT Example 15; Page 84-95; 167pp; English.
FT This sequence represents the nucleotide sequence of the second generation
FT adenovirus vector named Ad2-ORF6/PGK-CFTR. This virus lacks El and in its
FT place contains a modified transcription unit with the phosphoglycerate
FT kinase (PGK) promoter and a poly A addition site flanking the cystic
FT fibrosis transmembrane conductance regulator (CFTR) cDNA. The PGK
FT promoter is only if moderate strength but it is long lasting and is not
FT subject to shut off. The E4 region of the vector has also been modified
FT in that the whole sequence has been removed and replaced by ORF6, the
FT only E4 gene essential for growth of adenovirus (Ad) in tissue culture.
```

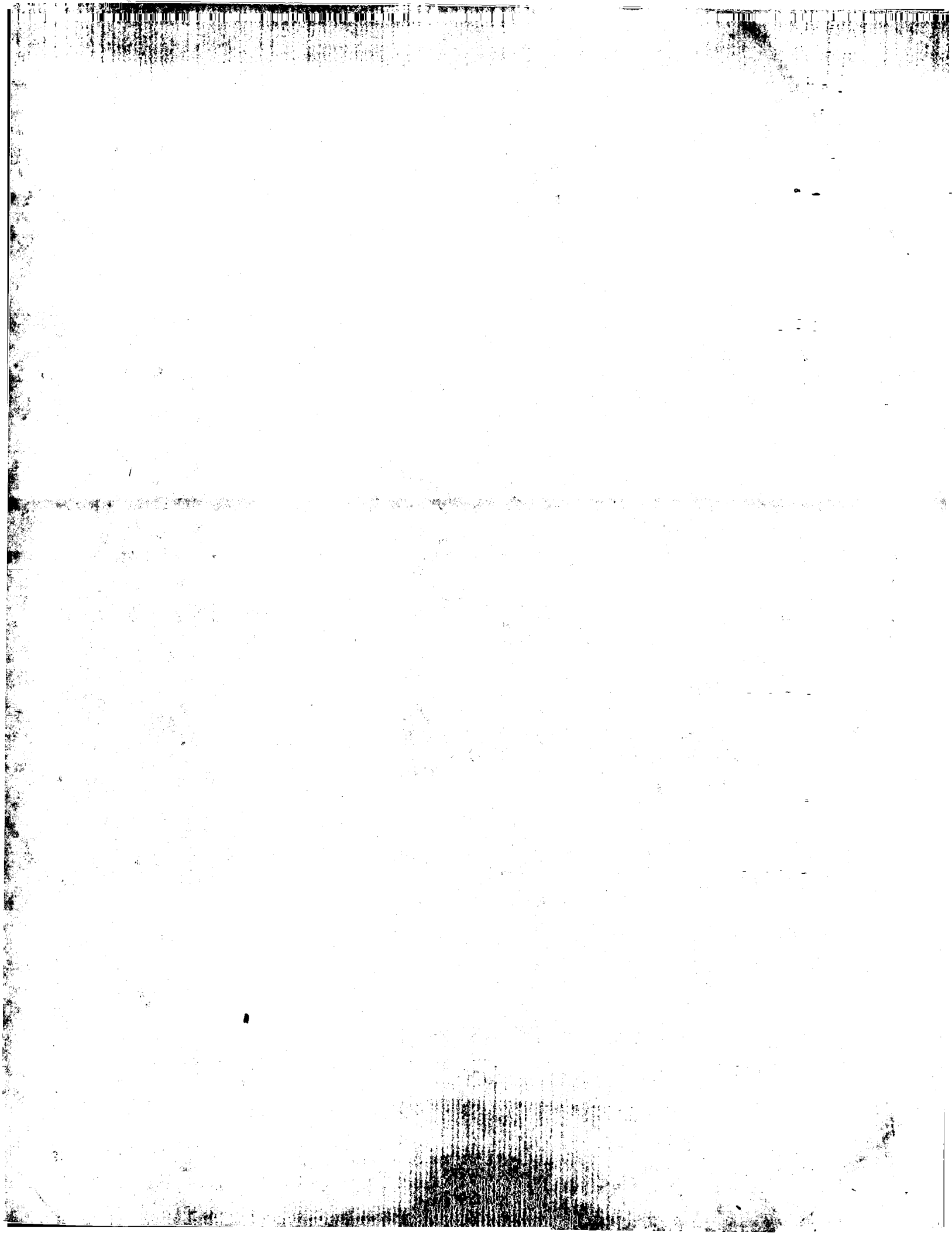

CC The DNA construct comprises a full length copy of the Ad2 genome from
CC which the early region 1 genes (E1 genes) have been deleted and replaced
CC by an expression cassette encoding CFTR. The expression cassette
CC includes the promoter for PGK and a poly-A addition signal from the
CC bovine growth hormone gene. The Ad2-ORF6/PGK-CFTR construct differs
CC from that given in Q68002 (Ad2/CFTR-1), in that the latter utilises the
CC endogenous E1a promoter, has no poly-A addition signal directly
CC downstream of CFTR and retains an intact E4 region. This adenovirus may
CC be administered to the pulmonary airways in the gene therapy of cystic
CC fibrosis.
SQ Sequence 36335 BP; 8597 A; 10000 C; 9786 G; 7952 T;

Query Match 98.9%; Score 875.4; DB 1; Length 36335;
Best Local Similarity 99.3%; Pred. No. 2.3e-265;
Matches 879; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 atgactacgtccggttcatttgcatgacactacgacacacacgatctcgggtgtct 60
Db 35968 ATGACTACGTCCGGCGTTCCATTGGCATGACACTACGACCAACACGATCTCGTGTCT 35909
QY 61 cggcgcactccgtacagtaggagctgtctacctctcttttgacagagaaacccgcgtacc 120
Db 35908 CGCGGCACCTCCGTACAGTAGGATCGCTACCTCTTTGACAGAGACCCGCGCTACC 35849
QY 121 atactggagatcatccgtgtgcccgaatgtaacactttgacaatgcacaaogtga 180
Db 35848 ATACTGGAGATCATCCGCTGTGCGCGAATGTAACACTTTGACAATGCACAACGTGAGT 35789
QY 181 tacgtgcgaggtctccctgcagtggtggattacgtgattcaggaatgggtgtctccc 240
Db 35788 TACGTGCGAGGTCTCCCTGCAGTGTGGATTACGCTGATTACGGAATGGGTGTTCCC 35729
QY 241 tgggataggttctaacgcggagagcttgaatccctgaggaagtgatgcacgtgtgc 300
Db 35728 TGGGATAGGTTCTGACGCGGGAGGAGCTGTGTAATCCTGAGGAAGTGTATGCAGTGTGC 35669
QY 301 ctgtgttggcacaatgatcatgacagcatgatccatggttaacagatcctgg 360
Db 35668 CTGTGTGTGCCAACATTGATATCATATGACGAGCATGATGATCCATGTTACGAGTCTCTG 35609
QY 361 gctctccactgtcattgtccagtcctccggttccctgcagtgatagccggcgaggtt 420
Db 35608 GCTCTCCACTGTCAATGTTCCAGTCCCGGTTCCCTGCAGTGCATAGCGGGGCGAGGTT 35549
QY 421 ttggccagctggttaggatggttggtggatggcgccatgtttaatcagaggtttatgg 480
Db 35548 TTGGCCAGCTGTTAGGATGGTGGATGGGATGGGCCCATGTTTAAATCAGAGGTTTATATGG 35489
QY 481 taccgggaggtggtgaattacaacatgccaaaagaggttaattgttccagcggttt 540
Db 35488 TACCGGAGGTGGTGAATTACAAATGCCAAAGAGGTAATGTTTATGTCACAGCGTGT 35429
QY 541 atgaggggtgcacacttaactacgtcgttggtagtgatggccacgtgggtttctgtg 600
Db 35428 ATGAGGGGTGCGCCCTTAATCTACCTCGCTTGTGTATGATGCCACGTGGGTTCTGTG 35369
QY 601 gtcccccgcattgagcttgatcacagccttcacactgtggattttgaacaatatgtg 660
Db 35368 GTCCCCGCCATGAGCTTGGATACAGCGCCCTTGACCTGTGGGATTTGAACAATATTGTG 35309
QY 661 gtgctgtcgtcagttactgtctgtatttaagttagatcaggtgctgctgtgcccgg 720
Db 35308 GTGCTGTGCTGCAGTACTGTCTGATTTAAGTGAGATCAGGGTGGCTGCTGTGCCCG 35249
QY 721 aggacaaggccttatgtcggggcggtgcgaatcatcgtcgtgagagaccactgccatg 780
Db 35248 AGGACAAGGCGTCTCAGTCGCGGGCGGTGCGAATCATCGCTGAGGAGACCATGCCATG 35189
QY 781 ttgtattcctgcaggagcgagcgggcgggcagcagttatttcgcgcgtgctgcagac 840
Db 35188 TTGTATTCTGCGAGGACGAGCGGCGGCGGCGGACGATTTATTCGCGCGCTGCTGCAGCAC 35129

QY 841 caccgcccctatctctgagtgacgacgattatgactctacccccatgtag 885
Db 35128 CACCGCCCTATCTCTGATGACGATTATGACTCTACCCCATGTAG 35084

Search completed: June 23, 2000, 10:05:42
Job time: 3610 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 23, 2000, 08:00:14 ; Search time 42.97 Seconds
(without alignments)
2677.141 Million cell updates/sec

Title: US-09-214-478-1
Perfect score: 885
Sequence: 1 atgactacgtccggtcc.....atgacttaccctccatgtag 885

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues
Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/1/ina/5A-COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B-COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/5C-COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/5D-COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/6-COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/PTUS-COMB.seq:*
- 7: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	885	100.0	8299	2	US-08-462-014-2
2	885	100.0	34303	4	US-08-735-609-4
3	885	100.0	34382	3	US-08-374-483-6
4	885	100.0	35935	4	US-08-735-609-1
5	885	100.0	35935	5	US-08-379-452-43
6	319.2	36.1	35081	3	US-08-752-760A-1
7	33.4	3.8	502	1	US-08-151-391A-3
8	33.4	3.8	509	1	US-08-151-391A-1
9	32.2	3.6	1391	4	US-08-865-336-2
10	30.6	3.5	7326	2	US-08-194-468-1
11	30.6	3.5	15894	1	US-08-348-891A-1
12	30.6	3.5	15894	2	US-08-905-817-1
13	30.4	3.4	972	6	PCT-US95-04801-2
14	30.4	3.4	4016	6	PCT-US95-08354A-1
15	30.4	3.4	6395	3	US-08-982-936-1
16	30.2	3.4	4378	4	US-09-080-897-3
17	30	3.4	9401	3	US-08-432-693-1
18	30	3.4	49272	2	US-08-614-770A-1
19	29.8	3.4	1866	5	US-08-909-742-1
20	29.2	3.3	906	4	US-08-656-906-24
21	29.2	3.3	5117	6	PCT-US95-05512-1
22	28.6	3.2	1941	5	US-09-082-737-1
23	28.6	3.2	4248	5	US-08-678-614-1
24	28.6	3.2	4757	1	US-07-865-662F-12
25	28.2	3.2	757	4	US-08-457-752-2
26	28.2	3.2	1317	1	US-08-228-162-1
27	28.2	3.2	2018	5	US-08-714-918-16

C 28	28.2	3.2	4181	1	US-07-670-611-1	Sequence 1, Appl
C 29	28.2	3.2	4181	1	US-08-220-674-1	Sequence 1, Appl
C 30	28.2	3.2	4181	1	US-08-445-186-1	Sequence 1, Appl
C 31	28.2	3.2	4181	2	US-08-446-549-1	Sequence 1, Appl
C 32	28.2	3.2	4181	3	US-08-446-550-1	Sequence 1, Appl
C 33	28.2	3.2	8082	2	US-08-306-691B-41	Sequence 41, Appl
C 34	28.2	3.2	8082	2	US-08-187-785-1	Sequence 1, Appl
C 35	28.2	3.2	8082	6	PCT-US93-06251-28	Sequence 28, Appl
C 36	28	3.2	37	2	US-08-591-492-16	Sequence 16, Appl
C 37	28	3.2	760	1	US-08-133-979A-20	Sequence 20, Appl
C 38	28	3.2	760	3	US-08-436-890-20	Sequence 20, Appl
C 39	28	3.2	760	3	US-08-451-213-20	Sequence 3, Appl
C 40	28	3.2	5738	1	US-08-409-995-3	Sequence 9, Appl
C 41	28	3.2	9401	6	PCT-US91-02225-9	Sequence 1, Appl
C 42	27.8	3.1	561	1	US-07-853-985A-1	Sequence 1, Appl
C 43	27.8	3.1	561	1	US-07-681-703B-1	Sequence 1, Appl
C 44	27.8	3.1	561	1	US-08-184-236-1	Sequence 1, Appl
C 45	27.8	3.1	561	3	US-08-407-410B-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-462-014-2
; Sequence 2, Application US/08462014
; Patent No. 5756283
; GENERAL INFORMATION:
; APPLICANT: Wilson, James M.
; APPLICANT: Fisher, Krishna J.
; APPLICANT: Gao, Guang-ping
; TITLE OF INVENTION: Method for Improved Production of
; TITLE OF INVENTION: Recombinant Adeno-Associated Viruses for Gene Therapy
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr., P.O. Box 457
; CITY: Spring House
; STATE: PA
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/462,014
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: UPNH1337USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8299 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; US-08-462-014-2

Query Match 100.0%; Score 885; DB 2; Length 8299;
Best Local Similarity 100.0%; Pred. No. 3.3e-282;
Matches 885; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 atgactacgtccggtccatttggcatcacactacgacacacacgattcgtgtgtct 60
|||||

STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States Of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/735,609
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02484
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 34303 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-08-735-609-4

Query Match	100.0%;	Score 885;	DB 4;	Length 34303;
Best Local Similarity	100.0%;	Pred. No. 7.1e-282;		

Matches	885	Conservative	0	Mismatches	0	Indels	0	Gaps	0
QY	1	atgactcagtcgcgcgttccatttggaatgacactacagcaacaacacgatactcgcggtgtct	60						
Db	32446	ATGACTAGTCGCGCGTTCCTATTTGGCATGACACTACGACCAACACGATCTCGGTGTCT	32887						
QY	61	cggcgactccgtacagtaggagatcgctcactcccttttgagacagaaacccggcgatacc	120						
Db	32386	CGGCGCACTCCGTACAGTAGGAGTCGTCTACCTCCTTTTGAGACAGAAACCCGCGCTACC	32327						
QY	121	atactgaggatcatccgtctgctgcgcgaatgtatacactttgacaattgcacaacgctgagt	180						
Db	32326	ATACTGGAGGATCATCCGCTGCTCCCGAATGTAACTTTGACAATGCACACGTGAGT	32267						
QY	181	tactgcagaggtctccctgcagctgagtgaggatttacgctgattcaggaaatgggtttgtccc	240						
Db	32266	TACGTGGAGGCTCTCCCTGCAGTGTGGGATTTACGCTGATTACGGAATGGTGTGTCCC	32207						
QY	241	tgggatatggtcttaacgcgcggaggagcttgtaatectcagggaagtgtatgcacgtgtgc	300						
Db	32206	TGGGATATGTTCTAACCGCGGAGGAGCTTGTAATCCTGAGGAAGTGTATGCACGTTGC	32147						
QY	301	ctgtgtgtgtccaaacttgatcatcgaagagatgatcatccatggttacagagctcgg	360						
Db	32146	CTGTGTGTGCCAACATTGATATCATGACGAGGATGATGATCATCGGTACGAGTCTCTGG	32087						
QY	361	gcctccactgtcaattgttccagctcccgttccctgcagctgataccgcgcggcgaggtt	420						
Db	32086	GCCTCCACTGTCTATTTGCCAGTCCCGGTCCCTGCAGTGTATACCCGCGGGCAGGTT	32027						
QY	421	ttggccagctggtttaggatggttggaatggcgcaatttaatacagagggtttatatgg	480						
Db	32026	TTGGCCAGCTGGTTTAGGATGGTGTGGATGGCGCATGTTTAATCAGAGGTTTATATGG	31967						
QY	481	taccggagaggttgatattacaacatgccaagaagaggttaatttatgtccacagctggtt	540						
Db	31966	TACCGGAGGTGTGTAATTACAACATGCCAAAGAGGTAATGTTTATGTCCAGCGTGTTT	31907						
QY	541	atgaggggtgcgcacttaatactacctgcgcttgtgtgattgatggccacgctgggttctgtg	600						

```

RESULT      2
US-08-735-609-4/c
; Sequence 4, Application US/08735609
; Patent No. 5955360
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Amalfitano, Andrea
; APPLICANT: Hauser, Michael A.
; APPLICANT: Kumar-Singh, Rajendra
; APPLICANT: Hartigan-O'Connor, Dennis J.
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP

```

Db 31906 ATGAGGCGTCCCACTTAATCTACCTCGCTTGGTATGATGCCACGCGGGTTCGTG 31847
Qy 601 gtccccccatgagcttgatagacagcgcccttgcaactgtgggatttgaacaataattgtg 660
Db 31846 GTCCCCGCAATGACCTTGGATACAGCGCCTTGACATGTGGGATTTTGAACAATATTGTG 31787
Qy 661 gtctgtgtcagcttactgtctgatttaagtgtgagatcaggggtgcctgtgtgcccgg 720
Db 31786 GTGCTGTGCTGCAGTACTGTGCTGATTTAAGTGTGAGATCAGGGTGCCTGCTGTGCCCGG 31727
Qy 721 aggaacaagcgcccttactgtcggcggtgcaaatcctcgtgagagacactgcccag 780
Db 31726 AGGACAAAGCGCCCTTATGCTCGGCGGTGGGAATCACTGCTGAGAGACACTGCCATG 31667
Qy 781 ttgtattctcgtcagagcagcgagcggtggcgagcagcttatttcgcgctgtcgtcagcac 840
Db 31666 TTGTATTCTCGACAGCGAGCGGCGGCGGAGCAGATTTATTTCGCGCGCTGCTGCAGCAC 31607
Qy 841 caccgcccctactcgtatgcagcagattactactaccccccatgtag 885
Db 31606 CACCGCCCTATCCTGATGACGATTTATGACTCTACCCCATGTAG 31562

RESULT 3
US-08-374-483-6/c
; Sequence 6, Application US/08374483
; Patent No. 5880102
; GENERAL INFORMATION:
; APPLICANT: GEORGE, SAMUEL E.
; APPLICANT: BLAZING, MICHAEL A.
; TITLE OF INVENTION: ADENOVIRAL VECTOR SYSTEM
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESS: NIXON & VANDERHYE P.C.
; STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,483
; FILING DATE: 17-JAN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 1579-83
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34382 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-374-483-6
Query Match 100.0%; Score 885; DB 3; Length 34382;
Best Local Similarity 100.0%; Pred. No. 7.1e-282;
Matches 885; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 atgactacgtcggcggttccatttggtgacactacacacacagctcgtgtctct 60

Db 32278 ATGACTACGTCCGGGTTCCTATTTGGCATGACACTACGACCAACACGATCTCGGTGTGCT 32219
Qy 61 cggcgcaactcgtacagtaggactctactcctcttttgagacagaaacccggcgctacc 120
Db 32218 CGGCGCACTCCCTGACAGTAGGATCTCTACTCTCTTTTGGAGACAGAAACCCGGCGCTACC 32159
Qy 121 atactggagatccctcgtcgtcccgaaatgaatacatttgacaatcacaaactgagt 180
Db 32158 ATACTGGAGGATCATCCGCTGTGCGCAATTAACACTTTGACAAATGCACAACGTGAGT 32099
Qy 181 tdcgtgaggggtctccctcgtcagtgaggatttaccgagattcaggaatgggtttctccc 240
Db 32098 TAGCTCGGAGGCTTCCCTCGAGTGTGGGATTTACGCTGATTCAGGAATGGGTTGTTCC 32039
Qy 241 tgggatatgttctaacgcgggagagcgttgtaactcctcgtgaggaagtgtatgcacgtg 300
Db 32038 TGGGATATGGTCTTAACCGCGGGAGGAGCTTGAATCCTCAGGAAGTGTATGCACGTGTC 31979
Qy 301 ctgtgttgccaacattgatcatcagcagcagcagcagcagcagcagcagcagcagcag 360
Db 31978 CTGTGTGTGCCAACATTGATATCATGACGAGCATGATCATGTTTACGAGTCTCTGG 31919
Qy 361 gctctccactgtcattgttccagctccggttccctcagctgtatagccgcgggcgaggtt 420
Db 31918 GCTCTCCACTGTCTATTGTTCCAGTCCCGGTTCCTCGAGTGTATAGCCGCGGCGCAGGTT 31859
Qy 421 ttggccagcgtggttaggaggttggtggatggcgccatgtttaatacagaggtttatattg 480
Db 31858 TTGGCCAGCTGGTTAGGATGGTGGTGGATGGCGCCATGTTTAAATCAGAGGTTTATATG 31799
Qy 481 taccgggaggtgtgaattacaacatgcacaaagagagagtaattgttaccagcgtgttt 540
Db 31798 TACCGGAGGTGTGTAATTAACAATGCCAAAGAGAGTAATGTTTATGTCACCGCTGTTT 31739
Qy 541 atgaggggtgcaccttaattcactcgtcgttgggtatgattgagcagcagtggttctgtg 600
Db 31738 ATGAGGGGTGCCACTTAATCTACCTCGCTTGTGGTATGATGCCACGCTGGTCTGTG 31679
Qy 601 gtccccccatgagcttggatgacagcgcccttgcactgtgggattttgacaataattgtg 660
Db 31678 GTCCCGCCATGAGCTTTGGATACAGCGCCTTGCACCTGTGGGATTTTGAACAATATTGTG 31619
Qy 661 gtctgtgtcagcttactgtctgatttaagtgtgagatcaggggtgcctgtgtgcccgg 720
Db 31618 GTCTGTGCTGCAGTACTGTGCTGATTTAAGTGTGAGATCAGGGTGCCTGCTGTGCCCG 31559
Qy 721 aggaacaagcgcccttactgtcgtcggcggtgcgaatcctcgtgagggagacactgcccag 780
Db 31558 AGGACAAGCGCCTTATGCTCGGCGGTGCGAATCATCGCTGAGAGACACTGCCATG 31499
Qy 781 ttgtattcctcagagcagcgagcggtggcgagcagcttatttcgcgctgctcagcac 840
Db 31498 TTGTATTCTTCAGGACGAGCGGCGGCGGAGGATTTATTCGCGCGCTGCTGCAGCAC 31439
Qy 841 caccgcccctactcgtatgcagcagattactactaccccccatgtag 885
Db 31438 CACCGCCCTATCCTGATGACGAGTATTACCTCTACCCCATGTAG 31394

RESULT 4
US-08-735-609-1/c
; Sequence 1, Application US/08735609
; Patent No. 5955360
; GENERAL INFORMATION:
; APPLICANT: Chamberlain, Jeffrey S.
; APPLICANT: Amalfitano, Andrea
; APPLICANT: Hauser, Michael A.
; APPLICANT: Kumar-Singh, Rajendra
; APPLICANT: Hartigan-O'Connor, Dennis J.
; TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP

Matches 885; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgactacgtccggttcattgcatgacactacgacacacgactcgtggtgtct 60
DB 34077 ATGACTACGTCGCGGCTTCATTGGCATGACACTACGACCAACAGACTCGGTGTCT 34018

QY 61 cggcgactccgtacagtaggagtcctacotccttttgagacagaaacccgcgtacc 120
DB 34017 CGGCGCACTCGGTACAGTAGGATCGTCTACCTCCCTTTGAGACAGAAACCCGCGTACC 33958

QY 121 atactgagatcatccgtcgtcccgatgtaacacttgacaaatgcacacgtgagt 180
DB 33957 ATACTGAGATCATCCGCTGCTGCCGAATGAACACTTTGACATGACACACGTGAGT 33898

QY 181 tacgtcgaggtctccctcgtcagtggtgggtttacgtcgtgattcaggaatgggtgtcc 240
DB 33897 TACGTGCGAGGTCTTCCTCGAGTGGGATTTACGCTGATTGAGGAATGGGTGTCTCC 33838

QY 241 cgggatatggtcttaacgcgggagagcttgtaactcgtgaggaagtgtacgcgtgtgc 300
DB 33837 TGGGATATGTTCTAACCGCGGAGAGCTTGTAATCCTGAGGAAGTGTATGCACGTGTC 33778

QY 301 ctgtgtgtgccaacatgatcatcacgagcatgatccatggtttacgagtcctgg 360
DB 33777 CTGTGTTGTGCCACATGATGATATGATGACGAGCATGATGATCCATGTTACGAGTCTGG 33718

QY 361 gctccactgtcattgttccagtcctcgtggttccctcagtgatagcgcggcgaggtt 420
DB 33717 GCTCTCCACTGTCTATTGTTCCAGTCCCGGTTCCCTGAGTGTATAGCGCGGCGCAGTT 33658

QY 421 ttggccagctggtttagatggtgtgtagtgcgcctggttttaacagaggtttatgg 480
DB 33657 TTGGCCAGCTGTTTGTAGATGTTGTGTGATGCGGCATGTTTATATCAGAGTTTATGG 33598

QY 481 taccggaggtgtgtaatacaacatgcacacagaggaaggttaattgttatgtccagcgtgtt 540
DB 33597 TACCGGAGGTGTGTAATTAACAACATGCCAAAGAGGTAATGTTATGTCAGAGGTGTT 33538

QY 541 atgaggggtgcacatgaatcatctacgtcgtgtgtgtatgtagtgcacgtggtttctgtg 600
DB 33537 ATGAGGGTGCACATTAATCTACCTGCGCTTGTGTGTATGATGCGCATGTTGTTCTGTG 33478

QY 601 gtcccgccatgactgtttgatacagcgcctgctgactgtgggattttgaacaattgtg 660
DB 33477 GTCCCGCCATGAGCTTGTGATACAGCGCTTGACATGAGGATTTGAACAATATTGTG 33418

QY 661 gtgctgtgtcagttactgtgtgtatttaagttagatcaggggtgcgtgtgtgtcccg 720
DB 33417 GTGCTGTGTGAGTACTGTGTGATTTAAGTAGATCAGGGTGTGTGTGCTGCTGCTG 33358

QY 721 aggaacaggccttattgtcggcggtgcgaatcatcgtgagagaccactgacctg 780
DB 33357 AGGACAGGCGCTTATGCTGCGGCGGTGCGATCATGCTGAGGAGACCATGCGCATG 33298

QY 781 ttgtattcctgcagcagcagcgcggcgagcagcagcagcagcagcagcagcagcagcagc 840
DB 33297 TTGTATTCTGACGAGGAGGAGCGGCGGCGCAGCAGTATTATCGCGCTGTGTGACGAC 33238

QY 841 caccgcctatcctgtatgcacgattatgactctacccccatgtag 885
DB 33237 CACCGCCCTATCTGATGACAGTATTGACTCTACCCCCATGTAG 33193

RESULT 6
US-08-752-760A-1/c
; Sequence 1, Application US/08752760A
; Patent No. 5877011
; GENERAL INFORMATION:
; APPLICANT: Armentano, Donna
; APPLICANT: Gregory, Richard J.
; APPLICANT: Smith, Alan E.
; TITLE OF INVENTION: CHIMERIC ADENOVIRAL VECTORS
; NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:
ADDRESSEE: Baker & Botts, L.L.P.
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
COUNTRY: U.S.A.
ZIP: 10112

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/752,760A
FILING DATE: 20-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Seide, Rochelle K
REGISTRATION NUMBER: 32,300
REFERENCE/DOCKET NUMBER: A31385
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-705-5000
TELEFAX: 212-705-5020
TELEX:

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 35081 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-752-760A-1

Query Match 36.1%; Score 319.2; DB 3; Length 35081;
Best Local Similarity 62.4%; Pred. No. 5.7e-95;
Matches 520; Conservative 0; Mismatches 308; Indels 6; Gaps 1;

QY 46 cgatctcgtgtgtctcgtggtcactccgtacagtaggagtcgtctacacctcttttgagaca 105
DB 33153 CAATCCAGCTCGTCCGCCATCATCATATCGCAGGCGCGTTTACCAGATCTGTATGAT 33094

QY 106 gaaacccgcgtaccactactcgtgaggtatccctcgtcgtcgtcgtcgtcgtcgtcgtcgt 165
DB 33093 GAAACCCGCTTCACTGACTGATCAACATCCCTCTGTGCCCCGATTTGTGTCATGAGAT 33034

QY 166 atgcacacgtgagttacgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 225
DB 33033 TATCTAATGTAGTTCGTGCTGCGGATACCATGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 32974

QY 226 gaatgggtgttccctcgtggtatggttctaacgcgggagaggttgaatccctgaggaag 285
DB 32973 GAGTTTCAGTCCCTGGGATATGATCCTCACCACAGAGAAATAAAATTTTAAAGA 32914

QY 286 tgtatgcagctgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 345
DB 32913 TGTATGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 32854

QY 346 ggttacgagctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 405
DB 32853 GGTACGAGCGCTGGATCTCCTGATTCACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 32794

QY 406 gccggcgaggttttggccagctggttttaggaggtggtggtggtggtggtggtggtggtggt 465
DB 32793 GCGGAGGACCGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 32734

QY 466 cagaggtttatgttaccgcggaggtggtggtggtggtggtggtggtggtggtggtggtggt 525
DB 32733 CAGCGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 32674

QY 526 atgtccagcgtgtttatgaggggtgcgccacttaactcactcgtcgttgggtatgatgggc 585

US-08-865-336-2

US-08-194-468-1

Qy	628	gccttgcactgtgggattttgaacaaatattgtggtgtgtgtcgcagttacttgtctgat	687
Db	6296	GCATTAGTGTGGGTTGATTAAAGATGTTTCAGCACCTGCTGCTGCTGAGGAGAG	623
Qy	688	ttaagtgaatcaggggtgcgtgcctgtgcccgaggacaaagcgctctatgctgcgg	744
Db	6236	CTGGGTGACTTTTAGGTCGCTAGCAGGCTCTTCACAGGCACTTTGGCAGATGCTCTG	6180

```
RESULT 11
US-08-348-891A-1
; Sequence 1, Application US/08348891A
; Patent No. 5654136
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Keiko
; APPLICANT: MORI, Takayuki
; APPLICANT: MAKINO, Satoshi
; TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,
; CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR
; TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/348,891A
; FILING DATE: 25-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,400
; FILING DATE: 10-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 3-293625
; FILING DATE: 14-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: KP-7501
; TELEPHONE: 703-521-2297
; TELEFAX: 703-685-0573
; TELEX: 248425 EMBON
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 108..1682
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1807..3327
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3438..4442
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 5458..7107
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 7271..9121
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9234..15782
; US-08-348-891A-1

Matches 81; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
Qy 91 cctccttttgagacagaaacccgcgtaccatactggagagatcatccgctgctccgaa 150
|||||
Db 2516 CCTCATTGGAAACGGAGATCGCGTCTTTATTGACAGGTGGTGCACCAATGCTCGAA 2575
Qy 151 tgtaacactttgacaatgcacaacgtgagttacgtgaggtcttccctgcagtgaggga 210
|||||
Db 2576 AGTCACTCGGAACCAACATCAGGGCCAGTGCACCTCGGGGAATGTCCTCGAGTATGTA 2635
Qy 211 ttacgctgattcagggaatgggtttccctgggatggttcta 255
|||||
Db 2636 GCAATGCCGCACTGATACAGAGTGGACACCCGAACTCTGGTACCA 2680

RESULT 12
US-08-905-817-1
; Sequence 1, Application US/08905817
; Patent No. 5824777
; GENERAL INFORMATION:
; APPLICANT: SASAKI, Keiko
; APPLICANT: MORI, Takayuki
; APPLICANT: MAKINO, Satoshi
; TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,
; CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR
; TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: YOUNG & THOMPSON
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,817
; FILING DATE: 04-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/348,891
; FILING DATE: 25-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/848,400
; FILING DATE: 10-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 3-293625
; FILING DATE: 14-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: 32,925
; REFERENCE/DOCKET NUMBER: KP-7501A
; TELEPHONE: 703-521-2297
; TELEFAX: 703-685-0573
; TELEX: 248425 EMBON
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15894 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 108..1682
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1807..3327
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3438..4442
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 5458..7107
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 7271..9121
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9234..15782
; US-08-905-817-1

Query Match 3.5%; Score 30.6; DB 1; Length 15894;
Best Local Similarity 49.1%; Pred. No. 8;
```

```

; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
PCT-US95-04801-2

Query Match          3.4%; Score 30.4; DB 6; Length 972;
Best Local Similarity 53.3%; Pred. No. 2;
Matches 64; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 699 cagggtgctgtgtgtgcccgaggacaaagcgcccttatgtcggcggtggaatcat 758
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 363 CAGCCTGTACCGCGCCGCGAGGAGACCGCGCGCTGCTGTGACCGCGCGGCACCTA 422
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 759 cgtgtaggagaccactgccatgttgtattcttcagagacagagcgcgcgagcaggtt 818
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 423 CGAGCGGAGGAGCTCGACACCCCTGCTCGACTGCGACCCGCTGTGCGCCTGCTT 482
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
PCT-US95-08354A-1/c
; Sequence 1, Application PC/TUS9508354A
; GENERAL INFORMATION:
; APPLICANT: Temple University - Of The
; APPLICANT: Commonwealth System of Higher Education
; TITLE OF INVENTION: JAK3 PROTEIN TYROSINE
; TITLE OF INVENTION: KINASE AND DNA ENCODING THE SAME
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seigel, Gonda, Lavorgna
; ADDRESS: 6 Monaco, P.C.

```

```

STATE: Pennsylvania
COUNTRY: U.S.A.
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb
COMPUTER: IBM PS/2
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/08354A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/272,368
FILING DATE: 8 July 1994
ATTORNEY/AGENT INFORMATION:
NAME: Monaco, Daniel A.
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 6056-203 PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-8383
TELEFAX: (215) 568-5549
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4016 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single stranded
TOPOLOGY: linear
PCT-US95-08354A-1

Query Match          3.4%; Score 30.4; DB 6; Length 4016;
Best Local Similarity 53.3%; Pred. No. 4.4;
Matches 64; Conservative 0; Mismatches 56; Indels 0; Gaps 0

QY 259 cgggaggagcttgtaatcctgcagggaagtgtatgcacgtgtgcctgtgttgcacaactt 318
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2304 CAGGAGGACTTCCGAGTCATGTGTCTCACCATCCACGACTCCCGCTCGGCCACGGAA 2245
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 319 gatatcatgacgacgatgatgatccattggtttcagagtcctggcgcttcacagtctattgt 378
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: June 23, 2000, 05:42:32 ; Search time 592.38 Seconds
(without alignments)
6055.408 Million cell updates/sec

Title: US-09-214-478-1
Perfect score: 885
Sequence: 1 atgactacgtcgcgcgttcc.....atgactatcccccatgtag 885

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues
Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: EST:
1: em_est1:
2: em_est2:
3: em_est3:
4: em_est4:
5: em_est5:
6: em_est6:
7: em_est7:
8: em_est8:
9: em_est9:
10: em_est10:
11: em_est11:
12: em_est12:
13: em_est13:
14: em_est14:
15: em_est15:
16: em_est16:
17: em_est17:
18: em_est18:
19: em_est19:
20: gb_est1:
21: gb_est2:
22: gb_est3:
23: gb_est4:
24: gb_est5:
25: gb_est6:
26: gb_est7:
27: gb_est8:
28: gb_est9:
29: gb_est10:
30: gb_est11:
31: gb_est12:
32: gb_est13:
33: gb_est14:
34: gb_est15:
35: gb_est16:
36: gb_est17:
37: gb_est18:
38: gb_est19:
39: gb_est20:
40: gb_est21:
41: gb_est22:
42: gb_est23:
43: gb_est24:
44: gb_est25:

45: gb_est26:
46: gb_est27:
47: gb_est28:
48: gb_est29:
49: gb_est30:
50: gb_est31:
51: gb_est32:
52: em_est20:
53: em_est21:
54: em_est22:
55: em_est23:
56: em_est24:
57: em_est25:
58: em_est26:
59: gb_est33:
60: gb_est34:
61: gb_est35:
62: gb_est36:
63: gb_est37:
64: gb_est38:
65: em_est27:
66: em_est28:
67: em_est29:
68: em_est30:
69: gb_est39:
70: gb_est40:
71: gb_est41:
72: gb_est42:
73: gb_est43:
74: gb_est44:
75: em_est31:
76: em_est32:
77: em_est33:
78: em_est34:
79: gb_est45:
80: gb_est46:
81: gb_est47:
82: gb_gss1:
83: gb_gss2:
84: gb_gss3:
85: gb_gss4:
86: em_gss1:
87: em_gss2:
88: em_gss3:
89: em_gss4:
90: gb_gss5:
91: gb_gss6:
92: gb_gss7:
93: gb_gss8:
94: gb_gss9:
95: em_gss5:
96: em_gss6:
97: em_gss7:
98: em_gss8:
99: em_gss9:
100: em_gss10:
101: em_gss11:
102: gb_gss10:
103: gb_gss11:
104: em_gss12:
105: gb_gss12:
106: gb_gss13:
107: gb_gss14:
108: gb_gss15:
109: gb_gss16:

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
C 1	521	58.9	564	79	AW265091	xp79d11.x
C 2	503	612	79	AW277220	AW277220	xq79b06.x
C 3	228.6	25.8	515	79	AW277227	xq80a02.x
C 4	228.6	25.8	515	81	AW419027	xy45f04.x
C 5	228.6	25.8	515	81	AW438779	xu38e04.x
C 6	218.8	24.7	493	80	AW340909	xy50e09.x
C 7	213.2	24.1	429	36	AA652034	ns48h01.s
C 8	201.4	22.8	301	36	AA649817	ns48f01.s
C 9	195.8	22.1	223	79	AW270676	xq56b11.x
C 10	194.6	22.0	445	36	AA642819	nu05b06.s
C 11	43.8	4.9	630	79	AW276836	xp66g11.x
C 12	42.2	4.8	922	82	CNS0073W	AL077673 Drosophila
C 13	40.4	4.6	848	82	CNS000GQ	AL066286 Drosophila
C 14	40.2	4.5	393	109	AO640309	AL072642 Drosophila
C 15	39.2	4.4	1101	82	CNS00LT2	AL078714 Drosophila
C 16	38.8	4.4	1101	82	CNS00KK2	AL106910 Drosophila
C 17	38.4	4.3	1101	83	CNS0161W	AL066286 Drosophila
C 18	38	4.3	895	82	CNS0071A	AL0648430 RPC193-Dp
C 19	37	4.2	424	109	AO648430	AO944567 Sheared D
C 20	37	4.2	537	92	AO944567	AO940148 Sheared D
C 21	37	4.2	544	92	AO940148	AO851204 LMAJFV1.1
C 22	36.6	4.1	511	91	AO851204	AL531303 SD02279.5
C 23	36	4.1	666	47	AT531303	AL059400 Drosophila
C 24	35.4	4.0	939	82	CNS00CNG	AL089964 HS_5304_A
C 25	35.2	4.0	881	91	AO899964	B12606 F24C4-T7 IG
C 26	35.2	4.0	827	84	B12606	AL106748 Drosophila
C 27	35.2	4.0	1101	83	CNS016HE	AL1596460 v17Ab01.x
C 28	34.8	3.9	434	48	AT1596460	AO647993 RPC193-Dp
C 29	34.8	3.9	442	109	AO647993	AO919044 RPC1-23-2
C 30	34.8	3.9	595	92	AO919044	AL065804 Drosophila
C 31	34.8	3.9	902	82	CNS006QP	AL019207 ub21a07.r
C 32	34.6	3.9	380	41	AI019207	AL063902 Drosophila
C 33	34.4	3.9	1101	82	CNS0038X	AA294303 SMOV3KCA2
C 34	34.2	3.9	400	31	AA294303	AO369543 HS_5033_B
C 35	34	3.8	600	106	AO369543	AL808482 wf08e12.x
C 36	34	3.8	791	60	AT808482	AL056736 Drosophila
C 37	34	3.8	1101	82	CNS00BBO	R72680 y193d03.r1
C 38	33.6	3.8	339	22	R72680	AA402831 zu49f10.r
C 39	33.6	3.8	415	33	AA402831	AA459525 zx89b09.r
C 40	33.6	3.8	429	34	AA459525	AL108707 Drosophila
C 41	33.6	3.8	1101	83	CNS017ZT	R30084 12669 lambda
C 42	33.4	3.8	511	23	R30084	AI994566 701498606
C 43	33.4	3.8	536	63	AI994566	AI660346 w662a08.x
C 44	33.2	3.8	399	49	AI660346	AI874404 tz65c07.x
C 45	33.2	3.8	436	61	AI874404	

ALIGNMENTS

RESULT 1
 LOCUS xp79d11.x1 NCI-CGAP_Ov40 Homo sapiens cDNA clone IMAGE:2746581 3' similar to SW:E434_ADE02 P03239 EARLY E4 34 KD PROTEIN. ;, mRNA
 DEFINITION
 ACCESSION AW265091 564 bp mRNA EST 28-DEC-1999
 VERSION xp79d11.x1 NCI-CGAP_Ov40 Homo sapiens cDNA clone IMAGE:2746581 3' similar to SW:E434_ADE02 P03239 EARLY E4 34 KD PROTEIN. ;, mRNA
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 564)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jul 7, 1999 this sequence version replaced gi:5406367.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Andrew Berchuck M.D., John Gillespie M.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: David B. Krizman, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: polyt not found
 Seq primer: -40UP from Gibco
 High quality sequence stop: 438.
 Location/Qualifiers

FEATURES

source
 1..564
 /db_xref="taxon:9606"
 /clone="IMAGE:2746581"
 /clone_lib="NCI-CGAP_Ov40"
 /sex="female"
 /tissue_type="endometrioid ovarian metastasis"
 /lab_host="DH10B"
 /note="Organ: ovary; Vector: pAMP10; cDNA made by oligo-dT priming. Non-directionally cloned into the UDG sites of pAMP10. Size-selected on agarose gel, average insert size 500 bp. Primary library, non-amplified. cDNA Library Preparation: David B. Krizman, Ph.D (NCI). Reference: Krizman et al. (1996) Cancer Research 56:5380-5383."
 BASE COUNT 158 a 175 c 117 g 112 t 2 others
 ORIGIN

Query Match 58.9%; Score 521; DB 79; Length 564;
 Best Local Similarity 98.7%; Pred. No. 3.3e-147;
 Matches 524; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 121 atactgagagatcatccgctgctgcccgaatgaacattgacattgacaaacagtgagt 180
 Db ATACTGNAGGATCATCGCTGCTCCGGAATGAACACTTGAACATGACAAACGAGT 505
 QY 181 taagtgcaggttccctcagtggtggtgattcacgtgtattcaggaattggtgtccc 240
 Db TAGTGCAGAGTCTTCCCTCAGTGTGGGATTACGCTGTATTACGGAATGGGTGTTCCN 445
 QY 241 tgggatatgttctaacgcggaggagcttgtaatacctcaggaagtgtatgcacgtgtgc 300
 Db TGGGATATGTTCTGACGGGGAGGAGCTGTGTAATCCTGAGGAAGTGTATGACGTGTGC 385
 QY 301 ctgtgttgcccaacattgatcatgacgagcatgatccatggttacgagtcctg 360
 Db CTGTGTTGTGCCAACATTGATATCATGACGAGCATGATCATCCATGTTACGAGTCTGG 325
 QY 361 gctctccatgtcatgttccagtcgccggttccctcagtcagtgatagccggcgagggtt 420
 Db GCTCTCCACTGTCTATTGTTCCAGTCCCGTCCCTCAGTGCATAGCCGCGGAGGTT 265
 QY 421 ttggccagctggttttaggatggtggtggtggtggtggtggtggtggtggttattatgg 480
 Db TTGGCCAGCTGTTTAGGATGTTGTTGATGGCGCCATGTTTAAACAGAGGTTTATATGG 205
 QY 481 taccggaggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 540
 Db TACCGGAGGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 145
 QY 541 atgaggggtgcgcacttaactacgtggtggtggtggtggtggtggtggtggtggtggt 600
 Db ATGAGGGGTGCGCACCTAATCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 85
 QY 601 gtcccgccatgagctgttggatgacagcgccttgctgactgtgggattttgaac 651
 Db GTCCCGCCATGAGCTTTGGATGACAGCCCTTGCACTGCTGGATTGTAAC 34


```

RESULT 2
AW277220/c
LOCUS
DEFINITION
sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 612)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
TITLE
Tumor Gene Index
JOURNAL
On May 19, 1998 this sequence version replaced gi:3137497.
COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
unknown library type
Possible reversed clone: polyT not found
Seq primer: -40UP from Gibco
High quality sequence stop: 375.
Location/Qualifiers
1..612
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2756819"
/clone_lib="NCI-CCGAP_lu34"
/tissue_type="large cell carcinoma"
/lab_host="DH10B"
/note="Organ: lung; Vector: pAMP10; cDNA made by oligo-
priming. Non-directionally cloned into the UDG sites o-
PAMP10. Size-selected on agarose gel, average insert
size 500 bp. Primary library; non-amplified. cDNA
Library Preparation: David B. Krizman Ph.D (NCI).
Reference: Krizman et al. (1996) Cancer Research
56:5380-5383."
BASE COUNT 206 a 185 c 93 g 124 t 4 others
ORIGIN
Query Match 56.8%; Score 503; DB 79; Length 612;
Best Local Similarity 91.5%; Pred. No. 9.9e-142; Indels 0; Gaps
Matches 530; Conservative
QY 73 tacagtaggtagctgtacctccctcttggagacagaaacccgcgtaccatactcgtggagat 132
Db 612 TATAGTAGGATCGTTAACTCTCTTTTGAGATAGAGATTTGTGTATCATATTTGGAGAGT 553
QY 133 catcgcgtctgcccgaatgaacacttgcacatgcacacgcagtcagtcgcaggt 192
Db 552 TATTGNTGTTGTCGAATGTAATCTTTGATGATGATATATGATGAGTATGTCGAGGT 493
QY 193 ctctccctcagctgggattacgcgtgattccaggaatgggtgttcctccgggatactggt 252
Db 492 TTTCNTCAGCTGGGGATTTATGCTGANTTAGGAATGGGTGTGTTCTTCGGGATATGGTT 433
QY 253 ctacacggggaggagctgttaactcctgaggaagtgtatcacgctgtgcctgtgttgcc 312
Db 432 CFCAGCTGGGAGGAGCTGTAACTCTTGAGGAGAGTGTATGTATGTCGTGTTGTGCC 373
QY 313 aacattgatcatcagcagcatgatccatccttggttacgagtcctcgtgcctcactgt 372
Db 372 AACATTGATATCATGATGAGCATGATCATCCATGGTTATGAGTTCTGGGCTTCACATGT 313
QY 373 cattgttcagctccgggttcctcagctgtatgccggcgggcaggttttggccagctgg 432
Db 432 cattgttcagctccgggttcctcagctgtatgccggcgggcaggttttggccagctgg 253

```


	768	gacaaactgcgaatgttattctcctgagacgagcgcgcgagcagtttatcttcgcgc	827
Qy			
	154	GACCACTGCCATCTTGATTCTTCGAGGACGAGCGGGCCGCGC-GCACGAGTATTTATCGGC	212
Db			
	828	gctgctgcagcacccacgcgcctatcctgatgcacgattatgactctaccgcccatatg	885
Qy			
	213	GCFTGTCGACACCACCGCCCTACTCTGATGCACGATTATGACTCTACGCCCATGTAG	270
Db			

RESULT 8
AA649817
LOCUS
DEFINITION
AA649817 301 bp mRNA EST 13-NOV-1997
ns48f01.s1 NCI-CGAP_A1v1 Homo sapiens cDNA clone IMAGE:1186873
similar to SW:E434_ADE02 P03239 EARLY E4 34 KD PROTEIN. ;, mRNA
sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Nov 29, 1993 this sequence version replaced gi:430266.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/abbr/image/image.html

```

Insert Length: 986      Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 275.
Location/Qualifiers
1. .301
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1186873"
/clone_lib="NCI_CGAP_Alvl"
/tissue_type="alveolar rhabdomyosarcoma"
/lab_host="DH10B"
/note="Vector: pAMPF10; mRNA made from alveolar
rhabdomyosarcoma, cDNA made by oligo-dT priming.
Non-directionally cloned. Size selected on agarose gel,
average insert size 600 bp. Reference: Krizman et al.
(1996) Cancer Research 56:5380-5383."
52 a 90 c 89 g 70 t

```

[illegible]

Db 156 CACTGCCATCTTATTCCTGCAGGACGAGCGCGC - GCACGAGTTATTTCGCGCT 214
 QY 831 gctgcagcaccacgcgcctatcctgatgcacattatgactaccccatag 885
 Db 215 GCTGCAGCACCGCGCCTATCCTGATGACAGNTAATGACTCTACCCCATGTAG 269

RESULT	9
AW270676	
LOCUS	xq56b11.x1 NCI CGAP Co22 Homo sapiens cDNA clone IMAGE:2754621 3'
DEFINITION	similar to SW:E434_ADE02 P03239 EARLY E4 34 KD PROTEIN.; mRNA sequence.
ACCESSION	AW270676
VERSION	AW270676
KEYWORDS	EST.
SOURCE	AW270676.1 GI:6657706 human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 223)
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	On Jun 5, 1998 this sequence version replaced gi:3189136. Contact: Robert Strausberg, Ph.D. Tel.: (301) 496-1550 Email: Robert_Strausberg@nih.gov unknown library type Trace considered overall poor quality Possible reversed clone: similarity on wrong strand Possible reversed clone: polyT not found Seq primer: -40UP from Gibco High quality sequence stop: 1. Location/Qualifiers 1..223
FEATURES	
source	

```

000000
1. 1223
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2754621"
/clone_lib="NCI CGAP_Co22"
/tissue_type="colonic adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pAMP10; cDNA made by oligo-dT
priming. Non-directionally cloned into the UDG sites of
pAMP10. Size-selected on agarose gel, average insert
size 500 bp. Primary library; non-amplified. cDNA
Library Preparation: David B. Krizman, Ph.D (NCI).
Reference: Krizman et al. (1996) Cancer Research
56:5380-5383."
38 a 56 c 76 g 53 t
ORIGIN
Query Match 22.1%; Score 195.8; DB 79; Length 223;
Best Local Similarity 92.4%; Pred. No. 7.6e-49;
Matches 206; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
Dy 660 ggctgtgtcgcagtactgtctgatttaagttagatcagggtgcctgtgtgcgg 719
bb 1 GGTCGTGCTGCAGTTACTGTGTGATTAAAGTAGATCAGGGTGCCTGCTGTGCCG 60
Dy 720 gaggacaaggcgcccttatgtctgcggcggtgcgaatcatcctaaggagacactgccat 779
bb 61 GAGGACAAAGGGCTCTATGCTGCGGGCGTCCGAATCATTCGTGAGGAGACACTGGCCT 120
Dy 780 gttgtattcttcgaggcgagcgcgcgccgagcagctttattcgcgcgtctgtcgacga 839
bb 121 GTGTGGTTCTGTGGGAACGAGCGGGGCGGAGCAGTTTATTGCGCGGCTGCTGCACGA 180
Dy 840 ccacgcgccctatctgtgcagcattatgactctaccacctg 882
bb

```

181 CCACGGCCCTATCTGATGCACGATTATGACTCTACCCCATG 223

RESULT 10
AA642819
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

27-OCT-1997
EST
1207091
IMAGE:
1207091
cdna clone
Homo sapiens
NCI-CGAP Al.v1
similar to SW:E417_ADE02 P03238 PROBABLE EARLY E4 17 KD PROTEIN ;
mRNA sequence.
AA642819
GI:2568037
AA642819.1
human.
Homo sapiens
Chordata; Craniata; Vertebrata; Mammalia;
Eukaryota; Metazoa; Chordata; Chordata; Chordata; Chordata; Chordata;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 445)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Sep 12, 1996 this sequence version replaced gi:1333370.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1350
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

```

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 339.
Location/Qualifiers
1. .445
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1207091"
/clone_1lb="NCL CGAP Alvl"
/tissue_type="alveolar rhabdomyosarcoma"
/lab_host="DH10B"
/note="vector: pAMP10; mRNA made from alveolar
rhabdomyosarcoma, cDNA made by oligo-dT priming.
Non-directionally cloned. Size selected on agarose gel,
average insert size 600 bp. Reference: Krizman et al.
(1996) Cancer Research 56:5380-5383."
BASE COUNT      94 a    114 c    132 g    105 t
ORIGIN

Query Match          22.0%; Score 194.6; DB 36; Length 445;
Best Local Similarity 94.9%; Pred. No. 2.5e-48;
Matches 223; Conservative 0; Mismatches 9; Indels 3; Gaps

QY 651 caatattgctgtcgtgcagttactgtctatttaagttagtcagggtcgctg 710
      |||||
DB   4 CCACATTGGTGCTGTCGTGCAGTTACTGCTGTTATTAAGTAGATCAGSGTCCGCTG 63

QY 711 ctgtgccggaggacaaggcgcccttatgtcgtggcggtgcgaataatcgtcgaggagac 770
      |||||
DB   64 CTGTGCC--GGAGACAAGCGCTCATGCTCGGGGGGTGCGAATCATGCTGAGAGAC 121

QY 771 cactgccaattgttattctcgtcaggaccgagcggcgcggcagcagttattcgcgcgt 830
      |||||
DB   122 CACTGCCATGTTGATTATCTCTCAGSAGCGAGCGCGGC-CACAGAGTTTATTCCGCCGCT 180

QY 831 gctgcagaccaccgcacctctctgatgcacgatttgactctacccccattgag 885
      |||||
DB   181 GTTGCAGCACACCACCGGCTATCTGTATGACGACGATTATGACTCPACCCCATGTAG 235

```

RESULT	11
AW276836	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

FEATURES SOURCE

BASE COUN-	ORIGIN
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

Query Match	Best Local Match
1	504

DB	584
RESULT	1
CNS0073W	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	

AW276836 630 bp mRNA EST 03-JAN-2000
 NCI-CGAP-Ov39 Homo sapiens cDNA clone IMAGE:2745380 3',
 similar to SW:E413_ADE02 P03240 PROBABLE EARLY E4 13 KD PROTEIN. ; ,
 mRNA sequence.
 AW276836
 AW276836.1 GI:5663866
 EST.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 630)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 On Nov 22, 1999 this sequence version replaced gi:56462296.
 Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Andrew Berchuck M.D., John Gillespie M.D.,
Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www.bio.linnl.gov/dbbrp/image.html

possible reversed clone: similarity on wrong strand
possible reversed clone: polyT not found
Seq primer: -40UP from Gibco
high quality sequence stop: 422.

Location/Qualifiers
 1. .630
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_xref="IMAGE:2745380"
 /clone_lib="NCI_CGAP_Ov39"
 /sex="female"
 /tissue_type="papillary serous ovarian metastasis"
 /lab_host="DH10B"
 /note="Organ: ovary; Vector: pAMP10; cDNA made by oligo-d
 priming. Non-directionally cloned into the UDG sites of
 pAMP10. Size-selected on agarose gel, average insert
 size 500 bp. Primary library; non-amplified. cDNA
 Library Preparation: David B. Krizman, Ph.D (NCI).
 Reference: Krizman et al. (1996) Cancer Research
 56:5380-5383."

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

```

      4.9%; Score 43.8; DB 79; Length 630;
      95.7%; Pred NO. 0.015;
      45: Conservative 0; Mismatches 2; Indels 0; Gaps
      atgactacgtcgcggcttcatttgcatgacactacgacaacag 47
      tcaatcattcgtccggcttcatttgcatgacactacgacaacag 630

```

CNSG0073W 922 bp DNA GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR14D09 of RPCr-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL066784
AL066784.1 GI:4945247

KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 922)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osogawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES source

BASE COUNT	223 a	95 c	TET3 ⁿ /note="end :	109 g	221 t	274 others
ORIGIN						

```

Query Match          4.8%; Score 42.2; DB 82; Length 922;
Best Local Similarity 16.9%; Pred. No. 0.055;
Matches 60; Conservative 132; Mismatches 162; Indels 0; Gaps 0;

```

[illegible][illegible]

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AL072642
AL072642.1 GI:4952523
GSS.
fruit fly.

Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 848)

REFERENCE	AUTHORS	TITLE	JOURNAL
1	Smith, J.	Effect of temperature on the rate of reaction of hydrogen peroxide with ferrous sulfate	J. Chem. Soc.
2	Johnson, A.	Study of the kinetics of the reaction between iodine and acetone	Trans. Faraday Soc.
3	Williams, B.	Investigation of the mechanism of the reaction of nitric oxide with carbon monoxide	Proc. Roy. Soc. (London)
4	Miller, C.	Analysis of the reaction between chlorine and hydrogen sulfide	J. Phys. Chem.
5	Thompson, D.	Experimental determination of the rate constants for the reaction of bromine with formic acid	J. Chem. Phys.
6	Roberts, E.	Thermodynamic study of the reaction between sulfur dioxide and carbon disulfide	J. Chem. Therm.
7	Clark, F.	Measurement of the activation energy for the reaction of ethyl acetate with hydroxide ions	J. Am. Chem. Soc.
8	White, G.	Reaction of hydrogen peroxide with various metal ions: a kinetic study	Can. J. Chem.
9	Green, H.	Reaction of hydrogen peroxide with ferrous ions in the presence of various catalysts	J. Chem. Soc. (A)
10	Black, I.	Reaction of hydrogen peroxide with ceric ions: a kinetic study	J. Chem. Soc. (B)
11	Gray, K.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (C)
12	Lee, M.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (D)
13	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (E)
14	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (F)
15	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (G)
16	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (H)
17	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (I)
18	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (J)
19	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (K)
20	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (L)
21	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (M)
22	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (N)
23	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (O)
24	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (P)
25	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (Q)
26	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (R)
27	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (S)
28	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (T)
29	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (U)
30	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (V)
31	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (W)
32	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (X)
33	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (Y)
34	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (Z)
35	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AA)
36	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AB)
37	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AC)
38	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AD)
39	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AE)
40	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AF)
41	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AG)
42	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AH)
43	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AI)
44	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AJ)
45	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AK)
46	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AL)
47	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AM)
48	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AN)
49	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AO)
50	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AP)
51	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AQ)
52	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AR)
53	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AS)
54	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AT)
55	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AU)
56	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AV)
57	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AW)
58	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AX)
59	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AY)
60	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (AZ)
61	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (BA)
62	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (BB)
63	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (BC)
64	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (BD)
65	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (BE)
66	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (BF)
67	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (BG)
68	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (BH)
69	Kim, S.	Reaction of hydrogen peroxide with ceric ions in the presence of various catalysts	J. Chem. Soc. (BI)
70			

TITLE Direct Submission
JOURNAL Submitted (02-JUN-

Genoscope (02.000.1393) Genoscope - Centre National de Séquençage
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Biosophila Genome Project (BBGP).
The snbp is constructed.

COMMENT

sequencing of this bac-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila* melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila* melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES	SOURCE
1. Feature 1	Source 1
2. Feature 2	Source 2
3. Feature 3	Source 3
4. Feature 4	Source 4
5. Feature 5	Source 5
6. Feature 6	Source 6
7. Feature 7	Source 7
8. Feature 8	Source 8
9. Feature 9	Source 9
10. Feature 10	Source 10
11. Feature 11	Source 11
12. Feature 12	Source 12
13. Feature 13	Source 13
14. Feature 14	Source 14
15. Feature 15	Source 15
16. Feature 16	Source 16
17. Feature 17	Source 17
18. Feature 18	Source 18
19. Feature 19	Source 19
20. Feature 20	Source 20
21. Feature 21	Source 21
22. Feature 22	Source 22
23. Feature 23	Source 23
24. Feature 24	Source 24
25. Feature 25	Source 25
26. Feature 26	Source 26
27. Feature 27	Source 27
28. Feature 28	Source 28
29. Feature 29	Source 29
30. Feature 30	Source 30
31. Feature 31	Source 31
32. Feature 32	Source 32
33. Feature 33	Source 33
34. Feature 34	Source 34
35. Feature 35	Source 35
36. Feature 36	Source 36
37. Feature 37	Source 37
38. Feature 38	Source 38
39. Feature 39	Source 39
40. Feature 40	Source 40
41. Feature 41	Source 41
42. Feature 42	Source 42
43. Feature 43	Source 43
44. Feature 44	Source 44
45. Feature 45	Source 45
46. Feature 46	Source 46
47. Feature 47	Source 47
48. Feature 48	Source 48
49. Feature 49	Source 49
50. Feature 50	Source 50
51. Feature 51	Source 51
52. Feature 52	Source 52
53. Feature 53	Source 53
54. Feature 54	Source 54
55. Feature 55	Source 55
56. Feature 56	Source 56
57. Feature 57	Source 57
58. Feature 58	Source 58
59. Feature 59	Source 59
60. Feature 60	Source 60
61. Feature 61	Source 61
62. Feature 62	Source 62
63. Feature 63	Source 63
64. Feature 64	Source 64
65. Feature 65	Source 65
66. Feature 66	Source 66
67. Feature 67	Source 67
68. Feature 68	Source 68
69. Feature 69	Source 69
70. Feature 70	Source 70
71. Feature 71	Source 71
72. Feature 72	Source 72
73. Feature 73	Source 73
74. Feature 74	Source 74
75. Feature 75	Source 75
76. Feature 76	Source 76
77. Feature 77	Source 77
78. Feature 78	Source 78
79. Feature 79	Source 79
80. Feature 80	Source 80
81. Feature 81	Source 81
82. Feature 82	Source 82
83. Feature 83	Source 83
84. Feature 84	Source 84
85. Feature 85	Source 85
86. Feature 86	Source 86
87. Feature 87	Source 87
88. Feature 88	Source 88
89. Feature 89	Source 89
90. Feature 90	Source 90
91. Feature 91	Source 91
92. Feature 92	Source 92
93. Feature 93	Source 93
94. Feature 94	Source 94
95. Feature 95	Source 95
96. Feature 96	Source 96
97. Feature 97	Source 97
98. Feature 98	Source 98
99. Feature 99	Source 99
100. Feature 100	Source 100

```

source
1. 848
/organism="Protophila melanogaster"
/db xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR3B02"
/notes="end : TTT3"

```

```

Query Match          4.6%; Score 40.4; DB 82; Length 848;
Best Local Similarity 22.6%; Pred. No. 0.18;
Matches 42; Conservative 74; Mismatches 70; Indels 0; Gaps 0;

```

375 ttgttccagtcgccggttccctgcagtgtagccggcgggcagggttttggccagctggtt 434

```
Db
703 KKTYSBGKKBTGKKKKKKTGKKKGDKGKGVKSGSGGGGKKGTGGGGGKCKGKK 644
```

435 taggatggtggtggaatggcgccatgttttaataatcagagggtttatatggtaccgggaggttgt 494

Db 643 GKDGGKXKGGARGGTTTTTKTKTKKTAWDWAAAAKKAATDTKKKKGTFRAGTGGRD 584

495 gaattacaacatgccaaaaagaggtaatgtttatgtccagcgtgtttatgaggggtcgcca 554

db 583 RAARKADGKAGKADGAADRRKRGKWTATTTTAKAKAWGSKTTTAAWKTCTKDWAATA 524

```
Qy 555 cttaat 560
    :: |:
pb 523 TWWKAK 518
```

RESULT 14

Accession	LOCUS	Size	Host	Strain	Date
AQ640309	LOCUS	393 bp	DNA	GSS	08-JUL-1999
	DEFINITION	927P1-4G2.TP 927P1 Trypanosoma brucei genomic clone 927P1-4G2, genomic survey sequence.			

ACCESSION	genomic survey sequenc
	AQ640309

VERSION AQ640309.1 GI:5117
KEYWORDS GSS

SOURCE Trypanosoma brucei.

ORGANISM
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.

Mon Jun 26 09:10:32 2000

us-09-214-478-1.rst

Page 10

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM-nucleic - nucleic search, using sw model

Run on: June 23, 2000, 09:34:53 ; Search time 878.41 Seconds

(without alignments)
-382.069 Million cell updates/sec

Title: US-09-214-478-3

Perfect score: 345

Sequence: 1 atgtttcttcagctcttc.....cggcgactcgtacagtag 345

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl:**

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_p11.*
- 8: gb_p12.*
- 9: gb_p1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_sy.*
- 14: gb_un.*
- 15: gb_v1.*
- 16: gb_v2.*
- 17: em_fun.*
- 18: em_hum1.*
- 19: em_hum2.*
- 20: em_in.*
- 21: em_om.*
- 22: em_or.*
- 23: em_ov.*
- 24: em_pat.*
- 25: em_ph.*
- 26: em_p1.*
- 27: em_ro.*
- 28: em_sy.*
- 29: em_un.*
- 30: em_v1.*
- 31: em_v2.*
- 32: gb_htg1.*
- 33: gb_htg2.*
- 34: gb_in1.*
- 35: gb_in2.*
- 36: gb_ba1.*
- 37: gb_ba2.*
- 38: gb_hum3.*
- 39: gb_hum4.*
- 40: gb_pr4.*
- 41: gb_htg3.*
- 42: gb_htg4.*
- 43: gb_htg5.*
- 44: gb_htg6.*

- 45: gb_htg7.*
- 46: em_htg1.*
- 47: em_htg2.*
- 48: em_htg3.*
- 49: em_hum5.*
- 50: gb_p13.*
- 51: gb_pr5.*
- 52: gb_htg8.*
- 53: gb_htg9.*
- 54: gb_htg10.*
- 55: gb_htg11.*
- 56: gb_htg12.*
- 57: gb_htg13.*
- 58: gb_htg14.*

Complement

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	345	100.0	3189	5	A52460	A52460 Sequence 4
C 2	345	100.0	35935	16	ADRCMPGEN	M73260 Mastadenovir
C 3	341.8	99.1	35937	16	ADRCG	J01917 Adenovirus
C 4	80	23.2	8299	5	AR009152	AR009152 Sequence
C 5	54	15.7	3119	16	S82508	S82508 984: Orf2..
C 6	54	15.7	35100	16	AF108105	AF108105 Human ade
C 7	52.8	15.3	3498	16	ADERB4	AF108105 Adenovirus
C 8	52.8	15.3	34125	16	AT12CGA	X73487 Adenovirus
C 9	36.8	10.7	34214	16	ADRGNOME	L19443 Human adenov
C 10	36.2	10.5	170245	10	HS109F14	AL022721 Human DNA
C 11	36	10.4	9971	12	NMGEPAD	X02801 Mouse gene
C 12	35.2	10.2	773	9	HUMSTXG1	M23945 Human chrom
C 13	35.2	10.2	2401	9	HUMSTB	J04964 Human stero
C 14	35.2	10.2	6520	9	HUMSTB	M15505 Human stero
C 15	35.2	10.2	99033	11	AC005704	AC005704 Homo sapi
C 16	35	10.1	486	3	AF075663	AF075663 Equus cab
C 17	35	10.1	14634	2	AE001442	AE001442 Helicobac
C 18	33.8	9.8	45410	41	AC006187	AC006187 Homo sapi
C 19	33.8	9.8	83336	55	AC022747	AC022747 Homo sapi
C 20	33.6	9.7	112629	8	F28D6	AF147242 Arabidops
C 21	33.4	9.7	211150	45	AC008732	AC008732 Homo sapi
C 22	33.2	9.6	6448	1	SCU58036	U68036 Streptomyce
C 23	32.6	9.4	214520	10	CNS0000J	AL049839 Human chr
C 24	32.6	9.4	278334	56	AC008543	AC008543 Homo sapi
C 25	32.4	9.4	40582	9	AP000691	AP000691 Homo sapi
C 26	32.2	9.3	872	40	AF112206	AF112206 Homo sapi
C 27	32.2	9.3	166180	32	CNS01DRF	AL117190 Homo sapi
C 28	32.2	9.3	184969	44	AC018362	AC018362 Homo sapi
C 29	32	9.3	4205	1	SGJ000264	AJ000264 Streptomy
C 30	31.8	9.2	771	13	G41969	G41969 SHGC-68997
C 31	31.8	9.2	35377	1	MTGY48	Z74020 Mycobacteri
C 32	31.8	9.2	112456	51	AC004614	AC004614 Homo sapi
C 33	31.8	9.2	216184	45	AC015899	AC015899 Homo sapi
C 34	31.8	9.2	225432	11	AF027390	AF027390 Homo sapi
C 35	31.6	9.2	188267	41	AC007734	AC007734 Homo sapi
C 36	31.4	9.1	1173	50	AF190094	AF190094 Magnolia
C 37	31.2	9.0	130690	1	D90901	D90901 Synchocyst
C 38	31	9.0	18746	42	AC015350	AC015350 Drosophil
C 39	31	9.0	118235	41	AC007854	AC007854 Drosophil
C 40	31	9.0	166536	56	AC008772	AC008772 Homo sapi
C 41	31	9.0	179127	41	AC007378	AC007378 Homo sapi
C 42	30.8	8.9	1735	11	HSU48361	U48361 Human NGFI-
C 43	30.8	8.9	1987	9	HSAD11081	AJ011081 Homo sapi
C 44	30.8	8.9	2192	10	HSNADR	X70591 H sapiens M
C 45	30.8	8.9	35698	2	AF213822	AF213822 Zymomonas

ALIGNMENTS

[illegible]

```

DGRGCEPRGSRFGRGRRVRRNTVTLGRPLAGAPEIGNRFRQHGYNLRSSGAAGTARSE
TOP"
26485..26514
/partial
/note="of a 33 K protein"
/number=1
26713..27083
/note="of a 33 K protein"
/number=2
27174..27857
/codon_start=1
/product="protein pVII"
/protein_id="AAA96413.1"
/db_xref="GI:454806"
/translation="MSKIPPTPYMWSYQPMGLAAGAAQDYSTRINYSAGPMHSIRV
NGTRAHNRILLEQAIAITTPRNINLRSPWPAALVYQESPAITYVLPRDAQAEYQMT
NSGAQLAGCFRHRVRSPOGITHLTIRGRGIQLNDESVSLSGLRDPDGTFFIQGAGRP
SFTPRQAAILTATSSERSRSGGICIGTIEEFVSFVFNPFSGPPGHYPDQFIIPNFDA
VKDSADGIV"
8367 a 10073 c 9761 g 7734 t
BASE COUNT
ORIGIN

```

```

Query Match      100.0%; Score 345; DB 16; Length 35935;
Best Local Similarity 100.0%; Pred. No. 6,6e-97;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgggtctccagctcttcccgctctcccggtgtgtgaactcgacagcaacaaatgttaggt 60
|||||
DB 34342 ATGGGTCTCTCCAGCTCTTCCCGCTCTCCCGTGTGTGACTCGCAGAACAAATGTGTAGT 34283
|||||

QY 61 tggctggggtgtggcttatctcgsgtgggtggatattatcaggcgacgcgcaetgaagga 120
|||||
DB 34282 TGCTGGGTGTGGCTTATTCTCGCGTGGTGGATCTTATCAGGCGACGCGCGCATGAGGA 34223
|||||

QY 121 gttacatagaacacgaacgacggggggcgctgcatgtctttgagagtggtatactac 180
|||||
DB 34222 GTTATCATAGAACCCGAGCCAGGGGGGCCCTGGATGCTTTGAGAGAGTGGATATACTAC 34163
|||||

QY 181 aactactacacagagcgatctaagcggcgagaccggagacgagatctgtttgtcacgcc 240
|||||
DB 34162 AACTACTATACAGAGCGATCTAAGCGGCGAGACCGGAGACGACGATCTGTTTGTCAAGCC 34103
|||||

QY 241 cgcacctggtttgtctcaggaaataactacgtccggcgcttcatttggcatgacact 300
|||||
DB 34102 CGCACCTGGTTTGTCTTACGAAATAATGACTACGTCCTCCGGCGCTTCATTGGCATGACACT 34043
|||||

QY 301 acgacaacacgactctcggttgtctcggcgacactccgtcacagtag 345
|||||
DB 34042 ACGACCAACACGATCTCGGTGTGCTCGGGCGCACTCCGTACAGTAG 33998
|||||

```

RESULT	3	ADRCG/c LOCUS	ADRCG	35937 bp	DNA	VRL	14-MAR-1996
DEFINITION		Adenovirus type 2, complete genome.					
ACCESSION		J01917	J01918	J01919	J01920	J01921	J01922
		J01926	J01927	J01928	J01929	J01930	J01931
		J01935	J01936	J01937	J01938	J01939	J01940
		J01944	J01945	J01946	J01947	J01948	J01949
		J01953	J01954	J01955	J01956	J01957	J01958
		M13004	V00007	V00008	V00009	V00010	V00011
		V00015	V00016	V00017	V00018	V00019	V00020
VERSION		J01917.1	G1:209811				
KEYWORDS		DNA polymerase; DNA-binding protein; RNA polymerase III; alternative splicing; coat protein; complete genome; genome-linked protein; glycoprotein; overlapping genes; polymerase; terminal repeat; unidentified reading frame; virus-associated RNA.					
SOURCE		Human adenovirus type 2.					
ORGANISM		Human adenovirus type 2					
REFERENCE		Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.					
AUTHORS		Ohe,K. and Weissman,S.M.					
		1 (sites)					

[illegible]

The nucleotide sequence of a low molecular weight ribonucleic acid from cells infected with adenovirus 2
J. Biol. Chem. 246 (22), 6991-7009 (1971)
72060437
2 (sites)
Jornvall, H., Ohlsson, H. and Philipson, L.
Biochem. Biophys. Res. Commun. 56 (2), 304-310 (1974)
74147480
3 (sites)
Celma, M.L., Pan, J. and Weissman, S.M.
Studies of low molecular weight RNA from cells infected with adenovirus 2. I. The sequences at the 3' end of VA-RNA I
J. Biol. Chem. 252 (24), 9032-9042 (1977)
78046048
4 (sites)
Celma, M.L., Pan, J. and Weissman, S.M.
Studies of low molecular weight RNA from cells infected with adenovirus 2. II. Heterogeneity at the 5' end of VA-RNA I
J. Biol. Chem. 252 (24), 9043-9046 (1977)
78046049
5 (bases 10514 to 10680)
Pan, J., Celma, M.L. and Weissman, S.M.
Studies of low molecular weight RNA from cells infected with adenovirus 2. III. The sequence of the promoter for VA-RNA I
J. Biol. Chem. 252 (24), 9047-9054 (1977)
78046050
6 (bases 30812 to 30900)
Zain, B.S. and Roberts, R.J.
Characterization and sequence analysis of a recombination site in the hybrid virus Ad2+ND
J. Mol. Biol. 120 (1), 13-31 (1978)
78153757
7 (sites)
Akusjarvi, G. and Pettersson, U.
Nucleotide sequence at the junction between the coding region of the adenovirus 2 hexon messenger RNA and its leader sequence
Proc. Natl. Acad. Sci. U.S.A. 75 (12), 5822-5826 (1978)
79096069
8 (bases 5986 to 6236)
Ziff, E.B. and Evans, R.M.
Coincidence of the promoter and capped 5' terminus of RNA from the adenovirus 2 major late transcription unit
Cell 15 (4), 1463-1475 (1978)
79084199
9 (bases 21607 to 21816)
Akusjarvi, G. and Pettersson, U.
Sequence analysis of adenovirus DNA. I. Nucleotide sequence at the carboxy-terminal end of the gene for adenovirus type 2 hexon
virology 91 (2), 477-480 (1978)
79119384
10 (bases 1 to 156; 35804 to 35937)
Arrand, J.R. and Roberts, R.J.
The nucleotide sequences at the termini of adenovirus-2 DNA
J. Mol. Biol. 128 (4), 577-594 (1979)
79175344
11 (sites)
Zain, S., Sambrook, J., Roberts, R.J., Keller, W., Fried, M. and Dunn, A.R.
Nucleotide sequence analysis of the leader segments in a cloned copy of adenovirus 2 fiber mRNA
Cell 16 (4), 851-861 (1979),
79211249
12 (bases 6039 to 6079; 7101 to 7172; 9634 to 9723; 18802 to 18861)
Akusjarvi, G. and Pettersson, U.
Sequence analysis of adenovirus DNA: complete nucleotide sequence of the spliced 5' noncoding region of adenovirus 2 hexon messenger RNA
Cell 16 (4), 841-850 (1979)
79211248
13 (sites)
Shinagawa, M. and Padmanabhan, R.
Nucleotide sequence at the inverted terminal repetition of

JOURNAL MEDLINE REFERENCE	adenovirus type 2 DNA Biochem. Biophys. Res. Commun. 87 (3), 671-678 (1979) 79209154
JOURNAL MEDLINE REFERENCE	14 (bases 25634 to 27376) Galibert,F., Herisse,J. and Courtois,G. Nucleotide sequence of the EcoRI-F fragment of adenovirus 2 genome Gene 6 (1), 1-22 (1979) 80004828
JOURNAL MEDLINE REFERENCE	15 (sites) Zain,B.S. and Roberts,R.J. Sequences from the beginning of the fiber messenger RNA of adenovirus-2 J. Mol. Biol. 131 (2), 341-352 (1979) 80029717
JOURNAL MEDLINE REFERENCE	16 (bases 5909 to 6178; 7023 to 7212; 9452 to 9836) Akusjarvi,G. and Pettersson,J. Sequence analysis of adenovirus DNA. IV. The genomic sequences encoding the common tripartite leader of late adenovirus messenger RNA J. Mol. Biol. 134 (1), 143-158 (1979) 80140449
JOURNAL MEDLINE REFERENCE	17 (sites) Perricaudet,M., Akusjarvi,G., Virtanen,A. and Pettersson,U. Structure of two spliced mRNAs from the transforming region of human subgroup C adenoviruses Nature 281 (5733), 694-696 (1979) 81012104
JOURNAL MEDLINE REFERENCE	18 (sites) Baker,C.C., Herisse,J., Courtois,G., Galibert,F. and Ziff,E. Messenger RNA for the Ad2 DNA binding protein: DNA sequences encoding the first leader and heterogeneity at the mRNA 5' end Cell 18 (2), 569-580 (1979) 80045037
JOURNAL MEDLINE REFERENCE	19 (sites) Zain,S., Gingeras,T.R., Bullock,P., Wong,G. and Gelinas,R.E. Determination and analysis of adenovirus-2 DNA sequences which may include signals for late messenger RNA processing J. Mol. Biol. 135 (2), 413-433 (1979) 80140500
JOURNAL MEDLINE REFERENCE	20 (bases 3504 to 4109) Alestrom,P., Akusjarvi,G., Perricaudet,M., Mathews,M.B., Klessig,D.F. and Pettersson,U. The gene for polypeptide IX of adenovirus type 2 and its unspliced messenger RNA Cell 19 (3), 671-681 (1980) 80155185
JOURNAL MEDLINE REFERENCE	21 (sites) Baker,C.C. and Ziff,E.B. Biogenesis, structures, and sites of encoding of the 5' termini of adenovirus-2 mRNAs Cold Spring Harb. Symp. Quant. Biol. 44 PT 1, 415-428 (1980) 81042935
JOURNAL MEDLINE REFERENCE	22 (bases 3536 to 35937) Shingawa,M., Padmanabhan,R.V. and Padmanabhan,R. The nucleotide sequence of the right-hand terminal SmaI-K fragment of adenovirus type 2 DNA Gene 9 (1-2), 99-114 (1980) 80202560
JOURNAL MEDLINE REFERENCE	23 (bases 27373 to 30050) Herisse,J., Courtois,G. and Galibert,F. Nucleotide sequence of the EcoRI D fragment of adenovirus 2 genome Nucleic Acids Res. 8 (10), 2173-2192 (1980) 81053687
JOURNAL MEDLINE REFERENCE	24 (sites) Akusjarvi,G., Mathews,M.B., Andersson,P., Vennstrom,B. and Pettersson,U. Structure of genes for virus-associated RNAI and RNAII of adenovirus type 2 Proc. Natl. Acad. Sci. U.S.A. 77 (5), 2424-2428 (1980) 80234635
JOURNAL MEDLINE REFERENCE	25 (sites) Perricaudet,M., Le Moullec,J.M. and Pettersson,U. Predicted structure of two adenovirus tumor antigens

us-09-214-478-3.rge

```

ITGDTKRAYSMOFLGASFGALKFSLTILASPTLSPGSELSAVVAQDLSDFLQLTLRRR
LRAEGRTLLNLVNTLQVVEQDLEFL"
1278. .1631
/gene="orf3"
1278. .1631
/gene="orf3"
/note="This sequence comes from Fig. 1"
/codon_start=1
/protein_id="AAB37506.1"
/db_xref="GI:1699393"
/translation="MKVCLLMKVEGALWELFMKCGVDLHQQFVEIIQGWKNENYLGWV
QECNLIDEIDGGFAENWILMDVRVPLELLEATVHLENRVGFEDLAVCFHQHSGGERL
HIROLHRIVLDRLE"
1634. .1996
/gene="orf4"
1634. .1996
/gene="orf4"
/note="This sequence comes from Fig. 1"
/codon_start=1
/protein_id="AAB37507.1"
/db_xref="GI:1699394"
/translation="MVLPLPPDLNDROGSINMGMAYRVLADVNRGIRMDGFFISS
DABELLQLNREWMTFSWMTPEQQKDRRRGCCSRAHFQKQYDKVRKRIHYNANRD
SIQLAPPSISQGPFTII"
1926. .2804
/gene="orf6"
1926. .2804
/gene="orf6"
/note="This sequence comes from Fig. 1"
/codon_start=1
/protein_id="AAB37508.1"
/db_xref="GI:1699395"
/translation="MOTRIQSSLRHHPYRRRLPRSEETRASLTBQHPLLPDCEHA
DHNVSVRGLSCAAGFILLQEPVPMDILTPBEIKILKRCMSVCLCPATLDLVRQAQ
MVSYGRIWLHCRSPGLSCQCTQGTLLSVFERYVYGCMFNOHFYWIROIYVNRNPP
KEIMYSGVFMGRHLIVICRIWDGVGSIIPNMSFGWSALNYGILLNNVIMCCTCYE
NMAIEMRCARBARTRLMLKAVGIIVRETDPDPDCSSRTIEPRQRLRLALMERHRTI
LFEEYSEVRSSSHSTRL"
2798. .3046
/gene="orf7"
2798. .3046
/gene="orf7"
/note="This sequence comes from Fig. 1"
/codon_start=1
/protein_id="AAB37509.1"
/db_xref="GI:1699396"
/translation="TVTLDCEARLDDFSEDGFSITDTPRLARQETWIIDPKSSRSTI
ENISLFKATRAERTVTVKAWGGRLTTBAGVKINKDT"
759 a 701 c 848 g 811 t
BASE COUNT
ORIGIN

Query Match 15.7%; Score 54; DB 16; Length 3119;
Best Local Similarity 52.28; Pred. No. 2.9e-06;
Matches 145; Conservative 0; Mismatches 130; Indels 3; Gaps 1

QY 1 atgggtcttcacagctctcccgctctcccgctgtgtgactcgcagacgaatgtgtaggt 60
Db 1634 ATGGTCTTCCAATCTGCCACCGCCCTCTGAATGATAGAC---AAGCGACGATTAA 1690

QY 61 tggctgggtgtggctattctgcggtgtgtgatatattcagggcgagcgcgatgaagga 120
Db 1691 TGGATGGGATGGCCTACAGATCTCGCTCATGTGATCAGGGGAATTCGATCGACGG 1750

QY 121 gttacatagacccggaacggggggcgctcggtgcttggagagagtgatatactac 180
Db 1751 TTTTATTTCATCATAGTCAGAGGAACCTCTCCAGAATCTTCAGAAATGGATGTACT 1810

QY 181 aactactacagagcgatctctaagcgcgcgacgcgagacgacgatctgttgtcacgc 240
Db 1811 AGTTGGATGACTGAGCGGCAGCAGCAAGAGGACGCCCGCAGAGGGGTATCTGCTGTCA 1870

```

```
QY 241 cgaactgtgttcttcaggaataatgactactgctcc 278
Db 1871 AGGGCCCACTTCTGCTGGCAGAAGTACGACAGGTACG 1908

RESULT 6
LOCUS AF108105 35100 bp DNA VRL 15-MAR-1999
DEFINITION Human adenovirus type 17 complete genome.
ACCESSION AF108105
VERSION AF108105.1 GI:4416335
KEYWORDS
SOURCE
ORGANISM Human adenovirus type 17.
Human adenovirus type 17
Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
REFERENCE 1 (bases 1 to 35100)
Chillon, M., Bosch, A., Zabner, J., Law, L., Armentano, D., Welsh, M. J.
and Davidson, B. L.
Group D adenoviruses infect primary central nervous system cells
more efficiently than those from group C
J. Virol. 73 (3), 2537-2540 (1999)
99139039
2 (bases 1 to 35100)
Armentano, D. PhD and Souza, D. W.
Direct Submission
Submitted (20-NOV-1998) Molecular Biology, Genzyme Corporation, 31
New York Ave., Framingham, MA 01701-9322, USA
LOCATION/Qualifiers
1. 35100
/organism="Human adenovirus type 17"
/db_xref="taxon:4922"
/complement(491..8266)
/codon_start=1
/product="DNA polymerase"
/protein_id="AAD20327.1"
/db_xref="GI:4416343"
/translation="MQCVDTATNSVPEIKYHLKHALTRLYEVNRLTPDLDLRT
MDSOLRALVAPLRRAEITWLPRLGIVLSVLEEGSHAGEHENHGPQLLKF
LKGRAVLVDVQVQRCCEGFRYKHQHCSSVRDRDFYFHHINSSHNWQIOFF
PISGPRERLEVTIDVETITWMSFGQLVFMVKFSGDPELVALARDLARLW
DWKSDPTFFCYTPERKAVQGFELFRDELQTLMAELWAFMQANPHQLQWALEOH
GLQCPEDTTEELKFLHPKIPKPRFMELYIVGHNIGDEIVLAAQVNNRASVPGF
RTNRMFRAGKILNDVTFALPNPLSKKRTDFELWEHGGCDDSDFKYQLKVMVRT
FALHTSLKAAQAVLPEVKGCCPKYAVNHFMVLSYRADDRGFPIRETKWDEETA
LNRELWKKGEAGVDLIRETDYCAMDVLTAEVAKLQDSIAHFIRDSVRLPHAFHN
IFORPTISNSHAFQIYVRAEQPTNFGGLGSPSHELIEYVRASIRGRCYPTN
IGISEPIYIDICMTASALTHPMPGPPPLNRYERAMAAREWQALDDSSKIDYFD
KLCGPIETIDAPDEHLDVLPFCRSKGGRLCWTNEPLRGVATSDVLTNLNRG
WRVIVDPKTTVPFWKCVAREYVQNLIAAKRADRKQNTWSTAKLLSNALYGSF
ATKLDNKIVFSDQMSLLKSIAAGQANIKSSSLETNLSAEMPALEREYLPQOL
ALVDSDAESEDHREPAPFTYTPSGTGHVAYTKPTTFDAEEDMCLTVEKVDPL
VNDNRYPSHVASFLVATRAVPSWSSEFLDEDRGTPLQDRPKSVIGDTSFLVTER
GHRLEWEGKRIKNGKLVDFPEQLTELWVECEVCAHCAADAFAPESVFLAPKL
YALASLCPAGRSKGLRAKHAALNMLNVCYLADAQGEDRAFSTSRMSLK
ATLASAOPGAPHPFTVTTTLTILRPWKMTLAALDAHLRVYPSRSRPNREVCWI
EMP"
complement(8311..10161)
/codon_start=1
/product="terminal protein"
/protein_id="AAD20326.1"
/db_xref="GI:4416342"
/translation="MDYFLPLRNWVRFFPRASITAAGITWMSRYLYGYHRLMED
LAPGAPATRWPLYRQPPHFLVGYOVLVNTCDYVDSRAFSRLRYSEVWQGLQTV
NWSLMANCTYITNGAVFVDMDDQDTLTRVQOAILAERVADALVQPLRGVGT
RMEDSASDIDRLMDHYKNLSRCQGWMAERLRCQAGPKDVLILLATRLKN
ATYNYLISNSNSVHRAATCLSLPCDDWDLAFERFSDVDLALTSPTQLIRCI
VSALSPLNGDPPHIREMTGGVTLRPRERGAFTETMRMRGEMIERFVRLPVRRR
RAPPPDPFFPAMEERLEAIGDISEMPLRWITFFVTEHTATLNTLYLQRLNVAFAE
HVELNLAQVYMRADADGVVYRWVNSGLGAFSOLMGRIENDLAATVERAGDGLQ
EEIEIQFMSIATQDNGSDVOETLRQPAVNDALIEISVLSFRFTYGPVTFQKQIQ
DVNRVVAHASALRAQHRLPERHADVPLPLPAGPEPPLPPGAPRHRH"
```

10622..11746
/codon_start=1
/product="L155K"
/protein_id="AAD20320.1"
/db_xref="GI:4416336"
/translation="MHPVLQMRPTTATTATAVAGAGASQPOQOTEMLEBEGGLA
RLGAPSPERHPVQLQKDVPAQYVPAQNLFRDRSGSEPEEMRDCFRFAGRELGLDR
QRVLDEDEFENEQIGISPARHVAANLVAVBOTVKQRNFQKFRAGNVRVRLIARE
YALGLMLDLAEAIQVONDSKPLTAQLFLVQHSRDNFAREALINIAEPGRWLL
ELINQISIVVOERSLSLAEKVAINYSLVSLACKFYARKYIKTPYPIDREKIDSYF
MRMAKVLITLSDDLGVYNDRIHKAVSASRRRLSRLSLRLVVGGAAGGESEY
FDGADLHWQPSRRALAAIGPEDLEDEEEEDAPAAGY"
11769..13463
/codon_start=1
/product="I11a"
/protein_id="AAD20321.1"
/db_xref="GI:4416337"
/translation="MQQAPDPAIRALQSPSGIASDDWEAMQRIALMTTRNPESFR
QQQANLSAILEAVPSRNPTRHEKVLAINALAEKNAIRPDEAGLVNALLRYGR
YNSVNSQNDRLVTDVREAVARERFKNEGLSLVALNAFLATQPAVPRGQDDYTR
FISALRLMTVEYQSEYQSGDYFTQSRQGLQTVNLQAFKRLGVOAQVQDGR
STVSSLTPNSRLILLIAPFTDSSGVNRSYGLHLLTYRAIGAOQVDEOTFORIT
SVSRALQNDTSLRATLNLNTRQOKIPAOYALSAEERILRYVQOQSVGLFMQEG
ATPSAALDMTARNMPSMYAANRPFINKLMDYLHRAASMSNYFTNAILNHLWLPFG
FYTGEXMDPNDGFLWDVDSVFTLQKQEAVRTPASEGAVGRSPFSLGSLHS
LPGSVNSGRVSRLLGDEDEYLNDLSLLQPPRVKNANNGIESLVDKLRNWKTYAQDH
REPAPRQRDRQRLGVWDDESDSSVLDLGGGVNPISRLQPKLGRMF"
13517..15070
/codon_start=1
/product="penton base"
/protein_id="AAD20322.1"
/db_xref="GI:4416338"
/translation="MRAVSSSPSPSVESVMAQATLEVFPVPRYMAPTEGRNSIRY
SELAPDITRVLYLVNDKADIASLNYQNDHNFLLTVQNNDFTPAASQTQINFDE
RSRWGGLKLTILHTNNVNYEYFTSKFARVMARKHPOGVEATDLKSDILEYEFER
FPLPEGSEFTMTIDLMNNAILEYLOVQNGVLESDIGVKFRSFRKLQWDPVKL
VMPVYVFAFDPVLLPGCGVDFTESRLSLGIRKQPOEQFRIMYEDLEGNNI
PALLDVKYLESKLLLEALEANAANGARGSDSVSREVEKAEKELVIEPIKODDT
KSVNLJEGMTDLYRSWLSYTRDENGVOGWTLLTPDVTGCAEQVWLSPLDMQ
DPYRNLQOQSVNYPVGAELMPPFRKSFINDLAVYSQILRSYTSLSLTFVFRPDNQI
LCRPAPITVTSYSENVPALMHTGLPRLSSIRGQVTVTDARRTCYVYKALGIVA
PVLSSRF"
17009..17713
/codon_start=1
/product="pVI"
/protein_id="AAD20323.1"
/db_xref="GI:4416339"
/translation="MEDINFASLAPRHGTFRFMGTWNEIGTSQNLGGAFNWSVWSGL
KNFSTLTYGNKAWNSSTGLLREKLQKQVQVQVGLASGVVYDIANAQVORE
INSRLDPPTTVEMEDATLPPKGEKRPDAEETILOVDPPSYEEAVKAGPTTR
IIAPLATGMKPAITLDLPPPTTAPPAAPVQPPVATVAVRVPARRQANWOSTLH
STVGLGVKSLARRCY"
20620..21249
/codon_start=1
/product="protease"
/protein_id="AAD20324.1"
/db_xref="GI:4416340"
/translation="MSGSSREELASIVRDLGCGTFWAPITQRPGLADKLCALAI
VNTAGRTGVHWAFAFGNPRSRICYMDFDPSDRRLKQIYSEYEAMLRSVAVSSP
DRCLSLQSGTQVGPSSAACGLFCMFLHAFVHWPDRPKQIYSEYEAMLRSVAVSSP
SPQVLTIRNQEELYRFLARHSYFRRSHRAAIEHATAFDKMLQRLVSYQ"
30935..32035
/codon_start=1
/product="fiber"
/protein_id="AAD20325.1"
/db_xref="GI:4416341"
/translation="MSKRVDEDFNVPYGYARNONIPFLTPPFVSSDGFKNFPFG
VLSKLADPITLIANGDVSGLVGLTQESMTVDKAPQLANNKLELVYDVEY
SANKLSKUPHGLIKLDKAGGLKDLIGKLVLTGKIGTENLQNTDSSRGIGTSV
RAREGTFDNGDYLAVANPKYDTRLTWTPTDTPNCRIDKDKSLTLVLTGQSGIL
ANSLVIVSGYQYIDHANTPTLKFIKLLFDNKGLVLPSSNLDSTYNFRSDNLT
SEAYKNAVEPMPLVAVPKPTGSKYARDIYGNILGLGLAYQPVVIVKIVTNEEDS
ASYTTEFVNKVEYAEVETSTFTSYIAQ"

CDS CDS CDS CDS CDS CDS CDS

J. Virol. 67 (2), 682-693 (1993)
 93124560
 (sites)
 REMARK
 MEDLINE
 4 (bases 4831 to 10470)
 Shu, L.M., Hong, J.S., Wei, Y.F. and Engler, J.A.
 Nucleotide sequence of the genes encoded in early region 2b of
 human adenovirus type 12
 Gene 46 (2-3), 187-195 (1986)
 87106854
 JOURNAL
 MEDLINE
 5 (bases 1 to 3957)
 van Ormondt, H. and Galibert, F.
 Nucleotide sequences of adenovirus DNAs
 Curr. Top. Microbiol. Immunol. 110, 73-142 (1984)
 85002829
 JOURNAL
 MEDLINE
 6 (bases 1 to 3957)
 Shinagawa, M. and Padmanabhan, R.
 Comparative sequence analysis of the inverted terminal repetitions
 from different adenoviruses
 Proc. Natl. Acad. Sci. U.S.A. 77 (7), 3831-3835 (1980)
 81054665
 JOURNAL
 MEDLINE
 7 (bases 1 to 3957)
 Sugisaki, H., Sugimoto, K., Takanami, M., Shiroki, K., Saito, I.,
 Shimojo, H., Sawada, Y., Uemizu, Y., Uesugi, S. and Fujinaga, K.
 Structure and gene organization in the transformed Hind III-G
 fragment of Ad12
 Cell 20 (3), 777-786 (1980)
 81022638
 JOURNAL
 MEDLINE
 8 (bases 1 to 3957)
 Tolun, A., Alestrom, P. and Pettersson, J.
 Sequence of inverted terminal repetitions from different
 adenoviruses: demonstration of conserved sequences and homology
 between SA7 termini and SV40 DNA
 Cell 17 (3), 705-713 (1979)
 80001962
 JOURNAL
 MEDLINE
 9 (bases 1488 to 3861)
 Kimura, T.
 Structure and sequence analysis of the transforming region E1B of
 human adenovirus type 12
 Sapporo Igaku Zasshi 52, 253-267 (1983)
 200966
 JOURNAL
 MEDLINE
 10 (bases 20966 to 22966)
 Kruijer, W., van Schaik, F.M., Speijer, J.G. and Sussenbach, J.S.
 Structure and function of adenovirus DNA binding protein:
 comparison of the amino acid sequences of the Ad5 and Ad12 proteins
 derived from the nucleotide sequence of the corresponding genes
 Virology 128 (1), 140-153 (1983)
 83277521
 JOURNAL
 MEDLINE
 11 (bases 24334 to 24703)
 Engler, J.A. and van Bree, M.P.
 The nucleotide sequence of the gene encoding protein IVa2 in human
 adenovirus type 7
 Gene 19 (1), 71-80 (1982)
 83054637
 JOURNAL
 MEDLINE
 12 (bases 1 to 530)
 Shibata, H., Zheng, J.H., Koikeda, S., Masamune, Y. and Nakanishi, Y.
 Cis- and trans-acting factors for transcription of the adenovirus
 12 E1a gene
 Blochm. Biophys. Acta 1007 (2), 184-191 (1989)
 89150250
 JOURNAL
 MEDLINE
 13 (bases 1 to 34125)
 Juttermann, R., Weyer, U. and Doerfler, W.
 Defect of adenovirus type 12 replication in hamster cells: absence
 of transcription of viral virus-associated and L1 RNAs
 J. Virol. 63 (8), 3535-3540 (1989)
 89311650
 JOURNAL
 MEDLINE
 14 (bases 1 to 34125)
 Kimura, T., Sawada, Y., Shinagawa, M., Shimizu, Y., Shiroki, K.,
 Shimojo, H., Sugisaki, H., Takanami, M., Uemizu, Y. and Fujinaga, K.
 Nucleotide sequence of the transforming early region E1b of
 adenovirus type 12 DNA: structure and gene organization, and
 comparison with those of adenovirus type 5 DNA
 Nucleic Acids Res. 9 (23), 6571-6589 (1981)
 82105565
 JOURNAL
 MEDLINE

REMARK (sites)
 FEATURES
 source
 repeat_unit
 variation
 variation
 promoter
 variation
 CDS
 CDS
 variation
 variation
 CDS

Location/Qualifiers
 1. .34125
 /organism="Human adenovirus type 12"
 /db_xref="taxon:28282"
 1. .161
 /function="inverted terminal repeat"
 6
 /citation=[5]
 /citation=[6]
 /citation=[7]
 /citation=[8]
 /citation=[12]
 /replace="a"
 8
 /citation=[5]
 /citation=[6]
 /citation=[7]
 /citation=[8]
 /citation=[12]
 /replace="t"
 9. .407
 /note="crossref Ad12 EPD30061 and EPD11197"
 /function="E1a promoter region (-499 to +100)"
 369
 /citation=[5]
 /citation=[6]
 /citation=[7]
 /citation=[8]
 /citation=[12]
 /replace="t"
 503. .1099
 /note="E1A; transcription activation; early protein;
 alternative splicing; crossref SWISS-PROT:E1A_ADE12,
 P03259"
 /codon_start=1
 /protein_id="CAA51877.1"
 /db_xref="GI:313362"
 /db_xref="SWISS-PROT:P03259"
 /translation="MRTMTPLVLSIQADLLEHLVNFNFNEVPSDDLLKVPVLYEL
 YLDVSEAGDNNEQVAEFPFESLILAASEGFLPEPVLSPVCEPIGSGCMPLQHP
 EDMDLCTEMGFPCSDSEDEQDGMHVSASAAAAADREEREFOLDHPELPGHNC
 SCEHRTGTNTDMLCSLCYLRYNMFYSCAMGGGR"
 1029. .1628
 /note="crossref Ad12 EPD07152"
 /function="E1b promoter region (-499 to +100)"
 1542. .2033
 /note="crossref SWISS-PROT:E1BS_ADE12, P04492"
 /codon_start=1
 /product="E1B protein, small T-antigen"
 /protein_id="CAA51878.1"
 /db_xref="GI:313363"
 /db_xref="SWISS-PROT:P04492"
 /translation="MELETVLQSFQSVRQLQYTSKNTSGFWYLFGLSTLSKVVRVK
 EDYREEFENILADCPGLLASLDLCYHLVQEKVVRSLDFSVSGVTASIAFLAILDK
 WSEKSHWSDYMLDYNMQLRAWLKRRCVCIYSLARPLTNPPLTLQEEEREENPAV
 VEK"
 1740
 /citation=[9]
 /replace="t"
 1742
 /citation=[9]
 /replace="g"
 1847. .3295
 /note="crossref SWISS-PROT:E1BL_ADE12, P04491"
 /codon_start=1
 /product="E1B protein, large T-antigen"
 /protein_id="CAA51879.1"
 /db_xref="GI:313364"
 /db_xref="SWISS-PROT:P04491"
 /translation="MEREIPELGLHAGLHVNAAVEGMAEEEGHLHLLAGAAFDHAAA
 DVARGEGGAGPCGGEVNMEOVQGHVLDSEGSFSCADDKDKKKSLKEAAVLS
 RLTVNLSRPRLETIVYQELQDEFQRGMHLQYKYSFEQLKTHWFEWDEMECAIKAF

23-NOV-1999

Human DNA sequence from clone 109F14 on chromosome 6p21.2-21.3. Contains the alternatively spliced gene for Transcriptional Enhancer Factor TEF-5, the 60S Ribosomal protein RPL10A gene, a PUTATIVE ZNF127 LIKE gene, and the PPARD for Peroxisome Proliferator Activated Receptor Delta (PPAR-Delta, PPAR-Beta, Nuclear Hormone Receptor 1, NUC1, NUCI, PPARB). Contains three putative CpG islands, ESTs, STSS, GSSs and a ca repeat polymorphism, complete sequence.

AL022721.1 GI:3367610
 HTG: 60S Ribosomal Protein RPL10A; ca repeat polymorphism; Cpg Island; NUC1; NUCI; Nuclear Hormone Receptor 1; Peroxisome Proliferator Activated Receptor Delta; PPAR-Beta; PPAR-Delta; PPARB; TEF-5; Transcriptional Enhancer Factor; ZNF127.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 170245)
 Tripodale,J., Mason,R., Humphray,S.J., Davies,A.F., Herberg,J.A., Townsend,J., Nizetic,D., Senger,G. and Ragoussis,J.
 Physical map of human 6p21.2-6p21.3: region flanking the centromeric end of the major histocompatibility complex
 Genome Res. 8 (6), 631-643 (1998)
 98311675
 2 (bases 1 to 170245)
 Phillips,S.
 Direct Submission
 Submitted (05-AUG-1998) E-mail enquiries: humquery@sanger.ac.uk
 Clone requests: clonerequest@sanger.ac.uk
 On Aug 1, 1998 this sequence version replaced gi:3355439.
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
 This sequence is the entire insert of clone 109F14. The true left end of clone 39A5 (597832) is at 133206 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
 This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed in collaboration by the Sanger Centre chromosome 6 mapping group and Ioannis Ragoussis. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6/>
 109F14 is from the library RPI1 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR.pCYPAC2>.
 Location/Qualifiers
 1. 170245
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /map="p21.2-21.3"
 /clone="RPI-109F14"
 /clone_lib="RPI-1"
 2790..2875
 /note="L1MB4 repeat: matches 842..927 of consensus"
 complement(6170..6295)
 /note="AluX repeat: matches 136..5 of consensus"
 6896..6893
 /note="MIR repeat: matches 33..231 of consensus"
 6872..6903
 /note="MIR2 repeat: matches 88..119 of consensus"
 7008..7342
 /note="AluSg repeat: matches 1..300 of consensus"
 complement(8324..8495)
 /note="AluSg repeat: matches 299..127 of consensus"

repeat_region
 repeat_region
 repeat_region
 repeat_region
 repeat_region
 repeat_region

FEATURES
 source
 /note="AluSg repeat: matches 300..3 of consensus"
 complement(8797..8925)
 /note="AluSg repeat: matches 129..1 of consensus"
 complement(9015..9183)
 /note="MIR repeat: matches 249..79 of consensus"
 join(9350..9603,11440..11504,15942..16004,17493..17630,17740..17789,18692..18753,19629..19762,19972..20145,20358..20498,20576..20728,20907..22465)
 /gene="dJ109F14.1"
 /note="match: cDNAs U46127 S74227 L06865 L13053 L26344 U51743 L26343 U63824 X94440 X94440 D87966 X94441 U10026 X94438 D87965 U06850 U06851 U04834 U06848 U06849; match: ESTs D87854 AAL150053 AA839812 AA779888 AA866660 AA270959 C18120 AAL156876 AAL150038 R81242 AA1308421 AA156787 W58359 A579824 AA587717 R27425 AA980487 AA972892 AA305329 W18351 W57598 AA613436 AA639669 AA534308 AA923822 W57598 AA613436 AA639669 AA534308 AA923822 AA852463 AA975080"
 /evidence=not_experimental
 /product="dJ109F14.1.2 (Transcriptional Enhancer Factor family protein)(PUTATIVE isoform 2)"
 join(9350..9603,11440..11504,15690..15752,16748..16759,17493..17630,17740..17789,18692..18753,19629..19762,19972..20145,20358..20498,20576..20728,20907..22465)
 /gene="dJ109F14.1"
 /note="match: cDNAs X94439 U46128 AF002670 D87964 Y16611 D87963 Y10027; match: ESTs D87854 AAL150053 AA839812 AA779888 AA866660 AA270959 C18120 AAL156876 AAL150038 R81242 AA308421 AA156787 W58359 AA579824 AA587717 R27425 AA980487 AA972892 AA305329 W18351 W57598 AA613436 AA639669 AA534308 AA923822 AA852462 AA852463 AA975080 W53780 W54739 AA797075"
 /evidence=not_experimental
 /product="dJ109F14.1.1 (Transcriptional Enhancer Factor TEF-5)(isoform 1)"
 9350..22465
 /gene="dJ109F14.1"
 join(<9405..9603,11440..11504,15942..16004,17493..17630,17740..17789,18692..18753,19629..19762,19972..20145,20358..20498,20576..20728,20907..21020)
 /gene="dJ109F14.1"
 /note="match: proteins Q90701 P28347 P30051 P48301 P48984 Q15561 Q62296 CE02703 P30052; the start codon used for the protein entries is not present in our sequence; we find an ata codon for isoleucine; the ORF continues upstream of this codon, into the next upstream exon where the true start could be located"
 /codon_start=1
 /evidence=not_experimental
 /product="dJ109F14.1.2 (Transcriptional Enhancer Factor family protein)(PUTATIVE isoform 2)"
 /protein_id="CAB38626.1"
 /db_xref="GI:4490421"
 /db_xref="SPTREMBL:O95910"
 /translation="ASNSWNASSPGAREDEGLDGLDNDAGVWSPDIEQSFOE ALAIYPCGRKIIISDECKMYRNEIARYIKLRTGTRKQVSSHQVLAARSKS EIOKLUQVSDKALQSNASSAQIYVSALQNKFPSPPLQAVFSTSSRFWSF PLLGQPGSPSDIKPQAFYFIQPLPPTLSSYEPLAPLSAASVPMODRTASS RLRLYSAFMEVQRPDPDTYSKHLFVHIGTNPALFSDPLEAVYQIYDFPEKKG LKELYKEPNAFFLVKFWADLNSTIQEGFAYGVSSOYSSADSKTISVTSKVGCSF KQVVEKVEYARLENGREVMYRIHRSPMCEYMINFIHLKLPKESMNVLENTEFII QVYTRSDSOETLLIVIAFVFEVSTSEHQAQHVYKLVKQ"
 join(<9405..9603,11440..11504,15690..15752,16748..16759,17493..17630,17740..17789,18692..18753,19629..19762,19972..20145,20358..20498,20576..20728,20907..21020)
 /gene="dJ109F14.1"
 /note="match: proteins P70210 Q99594; the start codon used for the protein entries is not present in our sequence; we find an ata codon for isoleucine; the ORF continues upstream of this codon, into the next upstream exon where the true start could be located"
 /codon_start=1
 /evidence=not_experimental

AUTHORS	Yen,P.H.
JOURNAL	Unpublished (1988)
COMMENT	There is a steroid sulfatase (STS) pseudogene on Yq.
FEATURES	Location/Qualifiers
source	1..6520
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/map="Xp22.32"
	<1..2429
sig_peptide	/note="steroid sulfatase mRNA"
	221..285
	/gene="STS"
	/note="steroid sulfatase signal peptide"
	221..1972
gene	/gene="STS"
	221..1972
CDS	/gene="STS"
	/note="steroid sulfatase precursor (EC 3.1.6.2)"
	/codon_start=1
	/db_xref="GDB:G00-120-393"
	/protein_id="AAA60596.1"
	/db_xref="GI:338514"
	/translation="MPLKGMKIPFLIFFLWEASHEASPNIILVMA YNGKITRPNDIRLASGGVKIHTOLAA SPLCTPSRAPMGRGPVRSG LTASSGLPIDEITPEALKIDQGYSTALIGKWHLGMSCKTDFCHH GISLNLRDCRPFGEVSVTTFGRKLVFPIQIVGVTLTLAALNCLGL LLFLAALLITPLGLFHFRPLNCFMNRNYEIQPNSYDNLTAQLTG ETFFLLLSYLHRVHTAFSSKDFAGKSQHVGDAVEMDWGSIVGILL TLITYTSOQGAHVVEVSFKGSIHGGSNGIYKGKANWEGGIRVGTLL KIDEPTSNMIDIFTVPVAKLAGAPLPEDRIIDGRDLMLPLEGKSORSH NAVRRHPONSSINKPAFTTFNFNPVNSGCFAITHVCFCFGSVTHHH PRERNPLTPASEPRFEYELIKVMQEAADRHQTQTUPEVPDQFSWNFLWK GLSCCCDREKODKRLSR"
mat_peptide	287..1969
	/gene="STS"
	/note="steroid sulfatase"
BASE COUNT	1881 a 1440 c 1363 g 1835 t
ORIGIN	Chromosome Xp22.32. 1 others

Query Match	10.26;	Score 35.2;	DB 9;	Length 6520;
Best Local Similarity	58.78;	Pred. No. 2.2;		
Matches	61;	Conservative 0;	Mismatches 43;	Indels 0;
Gaps	0;			
QY 239	ccgcgcactggtttgcttcaggaaaataactactgcgtccgcgcttcacttggcattgaca	298		
Db 936	CCTGCACTGCTTCATGATGAGGAACATACGAGATCATTCAGACGCCCTGCTCCTATGACA	995		
QY 299	ctacgacaaacacgactcgtgttctcggcgacactcgtctacag	342		
Db 996	ATCTCACCACGAGGCTAACGGTGGAGCGGCCCGCCAGTTCATACAG	1039		
RESULT 15				
AC005704/c				
LOCUS	AC005704	99033 bp	DNA	
DEFINITION	Homo sapiens Xp22 bins 29-30 BAC GSHB-227L7 (Genome Systems Human BAC Library) complete sequence.			26-SEP-1998
ACCESSION	AC005704			
VERSION	AC005704.1	GI:3659406		
KEYWORDS	HTG.			
SOURCE	Human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 99033) Muzny,D., Arenson,A.D., Adams,C., Brundage,E., Bunac,C., Carvelli,K., Chacko,J., Chen,J., Di,W., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Ganesh,R., Garcia,C., Goodman,M., Gorrell,J.H., Haywood,M., Hernandez,J., Jackson,L., Jin,S., Kampal,R., Karpaty,S., Kovar,C., Leal,B., Li,Y., Lichtarge,O., Liu,W., Logan,O., Lu,J., Ly,T., Martinez,C., Oswal,G., Perez,L., Rashid,N.D., Rowland,K., Savage,L., Scherer,S.E., Shen,H.,			

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2000, 09:07:32 ; Search time 20.4 seconds
(without alignments)
341.359 Million cell updates/sec

Title: US-09-214-478-2
Perfect score: 1594
Sequence: 1 WTSGVPGMTLRPTRSRLS.....RALLQHRPILMHLDYDSTPM 294

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1589	99.7	534	1 W03197	Glucocorticoid rec
2	1583	99.3	294	1 W01729	Human Ad5 E4 regio
3	101.5	6.4	383	1 W41592	Rat FRAG1 protein.
4	88.5	5.6	191	1 R37813	BamHI restriction
5	75.5	4.7	458	1 W48269	Open reading frame
6	74	4.6	1255	1 W01111	HER-2/neu protein.
7	74	4.6	1255	1 W92406	Human HER-2/neu on
8	73.5	4.6	265	1 W03549	ORF-5 protein sequ
9	73	4.6	391	1 W57324	Cytosolic glycerol
10	73	4.6	391	1 W60255	Klebsiella pneumonia
11	73	4.6	391	1 W30680	Glycerol-3-phospho
12	72	4.5	624	1 R08222	Extracellular port
13	71	4.5	369	1 W62647	Mature durum wheat
14	71	4.5	664	1 P60229	Alcohol-oxidase N
15	70.5	4.4	1058	1 R34843	HER4 with alternat
16	70.5	4.4	1058	1 R31734	Receptor tyrosine
17	70.5	4.4	1308	1 R54841	HER4. New recombin
18	70.5	4.4	1308	1 R31733	Receptor tyrosine
19	70	4.4	271	1 W82002	Human foetal brain
20	70	4.4	348	1 R38482	Rhodopsin protein.
21	70	4.4	348	1 R31116	Rhodopsin. Diagnos
22	70	4.4	664	1 R11606	Recombinant alcoho
23	70	4.4	782	1 W19764	Her2-QM-CSF immuno
24	70	4.4	2998	1 W19691	ATM mutant 3403del
25	70	4.4	3001	1 W19692	ATM mutant 4612del
26	70	4.4	3046	1 W19690	ATM mutant G9140T
27	70	4.4	3053	1 W19694	ATM mutant 7636del
28	70	4.4	3054	1 W19695	ATM mutant 7279del
29	70	4.4	3055	1 W19697	ATM mutant 5435del
30	70	4.4	3055	1 W19693	ATM mutant 8578del
31	70	4.4	3056	1 W19698	ATM mutant E2904G
32	70	4.4	3056	1 W19699	ATM mutant Leu43pr
33	70	4.4	3056	1 W19733	Cell cycle checkpo
34	70	4.4	3056	1 W37133	Ataxia-telangiecta

35 70 4.4 3056 1 W84269 Human ataxia telan
36 70 4.4 3059 1 W19696 ATM mutant 5319ins
37 70 4.4 3066 1 W36178 Murine Ataxia-tela
38 69.5 4.4 266 1 W06833 Polypeptide encode
39 69.5 4.4 662 1 R83304 Canine distemper v
40 69.5 4.4 855 1 R28033 Env polypeptide of
41 69.5 4.4 2963 1 W56444 Fragment HgJ1789 o
42 69 4.3 350 1 R68811 Interleukin-8 rece
43 69 4.3 350 1 R80756 Recombinant high a
44 69 4.3 350 1 R80951 ATM mutant C1240T.
45 69 4.3 413 1 W19682

ALIGNMENTS

RESULT 1

W03197
ID W03197 standard; Protein; 534 AA.
AC W03197;
DT 16-APR-1997 (first entry)
DE Glucocorticoid receptor/adenoviral ORF6 chimaeric fusion protein.
KW Human adenovirus; inverted tandem repeat; E4; E1; open reading frame;
KW expression; late gene; cell line; defective recombinant adenovirus;
KW complementation; gene therapy; cystic fibrosis; dystrophy; PCR; infection;
KW neurodegenerative disease; blood coagulation disorder; viral infection;
KW primer; amplification; polymerase chain reaction; fusion protein.
OS Homo sapiens/Adenovirus Ad5.
PN W09622378-Al.
PD 25-JUL-1996.
PF 19-JAN-1996; F00088.
PR 20-JAN-1995; FR-000747.
PR 01-JUN-1995; FR-006532.
PR 08-SEP-1995; FR-010541.
PA (RHON) RHONE POULENC RORER SA.
PI Dedieu J, Latta M, Orsini C, Perricaudet M, Vigne E;
PI Yeh P, Dedieu JF;
DR WPI; 96-354535/35.
DR N-PSDB; T31320.
PT Cells for prodn. of recombinant adeno and adeno-associated virus -
PT contain part of the E4 region of adeno virus, used to produce new
PT adeno virus defective in E1 and E4 regions for use as gene therapy
PT vectors
PS Example 1; Page 41-43; 69pp; French.
CC The invention relates to novel cell lines which can support the
CC replication and production of defective recombinant adenoviruses which
CC lack all or part of the E1 and E4 regions. The novel cell lines contain
CC the adenoviral E4 region, esp. the open reading frame 6 and/or 7,
CC inserted into its gene under control of a promoter, pref. from MTV which
CC is regulated by dexamethasone. The E4 ORF6 may also be inserted into the
CC genome as a chimaeric fusion gene comprising a glucocorticoid receptor
CC hormone binding domain fused N-terminally to the adenoviral ORF6 gene.
CC The fusion protein generated (this sequence) is retained in the cytoplasm
CC of the cell in the absence of hormone by interaction of the hormone
CC binding domain with the hsp90 protein or other factors. Contacting the
CC fusion protein with the hormone causes translocation of the protein to
CC the cell's nucleus where the ORF6 product functions. The gene was
CC generated by amplifying the sequence encoding the hormone binding domain
CC with primers T31316-7 and recombining the fragment with the adenoviral
CC ORF6 sequence. The defective viruses generated in the new cells can be
CC used for gene therapy e.g. for cystic fibrosis, dystrophy,
CC neurodegenerative diseases, blood coagulation disorders, viral
CC infections, etc.
SQ Sequence 534 AA;

Query Match 99.7%; Score 1589; DB 1; Length 534;
Best Local Similarity 100.0%; Pred. No. 2.6e-164;
Matches 293; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 TTSGVPGMTLRPTRSRLSRRTPYSRDLPPFETETATILEDHPLLPECTLTMHNSY 61
Db 242 TTSGVPGMTLRPTRSRLSRRTPYSRDLPPFETETATILEDHPLLPECTLTMHNSY 301

QY 62 YRGLPCSVGFTLLIQEWWVPMWVLTREELVILRKCMHVCLCCANIDIMTSMIHGYESWA 121
 DB 302 YRGLPCSVGFTLLIQEWWVPMWVLTREELVILRKCMHVCLCCANIDIMTSMIHGYESWA 361
 QY 122 LCHCSPGSLQIAGGQVLASWFRVVDGAMFNQRFIWRVNVNMPKVFMSVFM 181
 DB 362 LCHCSPGSLQIAGGQVLASWFRVVDGAMFNQRFIWRVNVNMPKVFMSVFM 421
 QY 182 RGRHLIYLRLWDGHSVVPAMSGYSAHCGILNNIVVLCSCYCADLSEIRVCCARR 241
 DB 422 RGRHLIYLRLWDGHSVVPAMSGYSAHCGILNNIVVLCSCYCADLSEIRVCCARR 481
 QY 242 TRRLMLRAVRIIAETAMLYSCRTERRRQOFIRALLQHRPILMHYDSTPM 294
 DB 482 TRRLMLRAVRIIAETAMLYSCRTERRRQOFIRALLQHRPILMHYDSTPM 534
 RESULT 2
 ID W01729 standard; Protein; 294 AA.
 AC W01729;
 DE 27-AUG-1997 (first entry)
 KW Human Ad5 E4 region ORF6 encoded by plasmid pMWV4ORF6 minigene.
 KW Minigene; human adenovirus type 5; open reading frame; promoter; in vivo;
 KW mouse mammary tumour virus; growth hormone; terminator; recombinant;
 KW packaging cell line; adeno-associated virus; ex vivo; gene therapy;
 KW inherited disease; cancer; genetic dysfunction; cystic fibrosis.
 OS Human adenovirus type 5.
 FH Key Location/Qualifiers
 FT misc_difference 115
 FT /note= "encoded by codon CAT"
 FT misc_difference 191
 FT /note= "encoded by codon TTG"
 FT W09639530-A2.
 PD 12-DEC-1996.
 PF 04-JUN-1996; U10245.
 PR 05-JUN-1995; US-462014.
 PR 27-OCT-1995; US-549489.
 FA (TYPE-) UNIV PENNSYLVANIA.
 PI Fisher KJ, Gao G, Wilson JM;
 DR WPI; 97-043152/04.
 DR N-PSDB; T59270.
 PR Recombinant adeno-associated virus contg. second gene which
 PT facilitates its conversion from single stranded to double stranded
 PT virus - enhances efficiency of ex vivo transduction into target cell
 PS Example 1; Page 72-73; 131pp; English.
 CC This is the amino acid sequence of the product of the human adenovirus
 CC type 5 (Ad5) E4 region open reading frame (ORF) 6. The coding sequence
 CC was placed under control of the mouse mammary tumour virus promoter and
 CC a growth hormone terminator sequence and used to create the minigene in
 CC plasmid pMWV4ORF6. The minigene is used to generate a packaging cell
 CC line expressing the E4 ORF6 product which is used to rescue and package
 CC an E4 deficient recombinant adeno-associated virus (rAAV). The rAAV
 CC comprises at least 2 selected genes, the first gene able to express the
 CC gene of interest in the target cell; and the second gene able to convert
 CC the single stranded virus to its double stranded form upon expression.
 CC The rAAV is useful in pharmaceutical compositions for use in ex vivo and
 CC in vivo gene therapy for the treatment of inherited diseases, cancer and
 CC other genetic dysfunctions, e.g. cystic fibrosis.
 SQ Sequence 294 AA;

Query Match 99.3%; Score 1583; DB 1; Length 294;
 Best Local Similarity 99.3%; Pred. No. 5.1e-164;
 Matches 292; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MTTSGVFGMTLRTSRSLRRTPYSRDLPPPTETRATILEDHPLPCNTLTHNVS 60
 DB 1 MTTSGVFGMTLRTSRSLRRTPYSRDLPPPTETRATILEDHPLPCNTLTHNVS 60
 QY 61 YRGLPCSVGFTLLIQEWWVPMWVLTREELVILRKCMHVCLCCANIDIMTSMIHGYESW 120

DB 61 YRGLPCSVGFTLLIQEWWVPMWVLTREELVILRKCMHVCLCCANIDIMTSMIHGYESW 120
 QY 121 ALHCHCSPGSLQIAGGQVLASWFRVVDGAMFNQRFIWRVNVNMPKVFMSVFM 180
 DB 121 ALHCHCSPGSLQIAGGQVLASWFRVVDGAMFNQRFIWRVNVNMPKVFMSVFM 180
 QY 181 MRGRHLIYLRLWDGHSVVPAMSGYSAHCGILNNIVVLCSCYCADLSEIRVCCARR 240
 DB 181 MRGRHLIYLRLWDGHSVVPAMSGYSAHCGILNNIVVLCSCYCADLSEIRVCCARR 240
 QY 241 TRRLMLRAVRIIAETAMLYSCRTERRRQOFIRALLQHRPILMHYDSTPM 294
 DB 241 TRRLMLRAVRIIAETAMLYSCRTERRRQOFIRALLQHRPILMHYDSTPM 294
 RESULT 3
 ID W41592 standard; Protein; 383 AA.
 AC W41592;
 DE 18-JUN-1998 (first entry)
 KW Rat FRAG1 protein.
 KW Fibroblast growth factor receptor activating gene 1; FRAG1; rat; FGFR2;
 KW fibroblast growth factor receptor 2; transforming activity; osteosarcoma;
 KW autophosphorylation activity; chromosomal abnormality detection;
 KW neoplasia.
 OS Rattus sp.
 FH Key Location/Qualifiers
 FT Peptide 1..129
 FT /note= "signal peptide"
 FT Protein 130..383
 FT /note= "mature FRAG1"
 FT Misc_difference 185
 FT /note= "encoded by GTG"
 FT Misc_difference 245
 FT /note= "encoded by CAC"
 FT W09748813-A2.
 PD 24-DEC-1997.
 PF 18-JUN-1997; U10660.
 PR 18-JUN-1996; US-020009.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Lorenzi MV, Miki T;
 DR WPI; 98-063151/06.
 DR N-PSDB; V04269.
 PT Rat and human FRAG1 nucleic acid and related proteins - used for
 PT diagnosis of neoplasia, particularly osteosarcoma
 PS Claim 13; Fig 4; 35pp; English.
 CC This sequence is the rat fibroblast growth factor receptor activating
 CC gene 1 (FRAG1) protein of the invention. When expressed in frame
 CC with FGFR2 (fibroblast growth factor receptor 2), the FRAG1 gene
 CC stimulates the transforming (autophosphorylation) activity of FGFR2. The
 CC FRAG1 gene can be used as a hybridisation probe to detect chromosomal
 CC abnormalities, particularly a rearrangement that fuses FRAG1 to another
 CC gene, resulting in expression of a fusion protein. Detection of such
 CC rearrangements is used to diagnose neoplasia, specifically osteosarcoma.
 CC These probes (or related primers) can also be used to identify related
 CC genes in other species. Agents that inhibit FRAG1 (e.g. antibodies or
 CC antisense molecules) may be useful therapeutically and the FRAG1 protein
 CC can also be used to screen for drugs. Antibodies are used for
 CC purification of the FRAG1 protein; in cloning FRAG1 alleles and
 CC homologues from expression libraries, and as probes in protein blots and
 CC immunoassays.
 SQ Sequence 383 AA;

Query Match 6.4%; Score 101.5; DB 1; Length 383;
 Best Local Similarity 22.2%; Pred. No. 0.0057;
 Matches 66; Conservative 36; Mismatches 106; Indels 89; Gaps 17;
 QY 11 TLRPTRSLRRTPYSRDLPPPTETRATILEDHPLPCNTLTHNVS 56
 DB 40 SLRPTPERPRLTPLLRLPAPSPGWAGAPGAPLATGVPPP--SHYLGPSAQDPSR 97
 QY 57 HNVSVYRG-----LPC---SVGFTLLIQEWWVPMWVLTREELVILRKCM--HV 99

Db 98 H-CGWRGGALEPQTSFSLPCQSLACGFESDKMYQVP--LTLDKRGTVLRLFTWVALI 154
QY 100 CLCCANID-----IMTSMIHGYSWALHCHSCSPGSLQCIA---GGQVLASWFRWVDGA 152
Db 155 TVCCPLVAFFCIWLSLLFHFKETSTHCGC--PNYLPSSVATGGEV----- 201
QY 153 MFORFIWYREVYNNMPKEYMFMSSVPMRGRHLIYLRWDGHVGVVPPAMSGYSAL- 212
Db 201 --PQRYVW-----RFGICGLHSAPELTAFAFW-NHYLSCASPCP--GYRLLC 242
QY 212 ----HCGILNIVVLCSSYCADLSEIRVRCARRTRMLRAVRIIAETAMLYSC 264
Db 243 RLNFSLNVVENLALLVLYVSSSEDFTHENA-----FVFIASLSYMLITC 290

RESULT 4

R37813 ID R37813 standard; Protein; 191 AA.
AC R37813;
DT 13-OCT-1993 (first entry)
DE BamHI restriction fragment 1 (BHRFI) gene product.
KW Antisense oligonucleotide; hybridisation; EBV;
KW Infection; infectious mononucleosis; Hodgkin's disease;
KW Burkitt's lymphoma; undifferentiated nasopharyngeal carcinoma;
KW AIDS related lymphoma; angioimmunoblastic lymphadenopathy;
KW EBV-induced proliferative disease; bone marrow.
OS Epstein-Barr virus.
PN W09311267-A.
PD 10-JUN-1993.
PF 04-DEC-1992; G02256.
PR 05-DEC-1991; GB-025891.
PA (CANC-) CANCER RES INST.
PI Cunningham D, Hickish T;
DR WPI: 93-197083/24.
DR N-PSDB: Q39619.
PT New anti-sense oligo-nucleotide which hybridises with BHRFI gene of EBV - for treating Epstein-Barr virus infection, infectious mononucleosis, Hodgkin's disease, Burkitt's lymphoma etc.
PS Disclosure; Page 11-12; 20pp; English.
CC The oligonucleotides given in Q39617-18 are selectively hybridisable to the BHRFI gene (Q39619). The antisense oligonucleotides can be used to treat EBV infection in patients suffering from infectious mononucleosis, Hodgkin's disease, Burkitt's lymphoma, undifferentiated nasopharyngeal carcinoma, AIDS related lymphoma or angioimmunoblastic lymphadenopathy. They can also be used to treat EBV-induced proliferative diseases by treating bone marrow from a patient to destroy proliferating tumour cells, then reimplanting the marrow into the patient.
SQ Sequence 191 AA;

Query Match 5.68; Score 88.5; DB 1; Length 191;
Best Local Similarity 23.38; Pred. No. 0.038;
Matches 49; Conservative 18; Mismatches 66; Indels 77; Gaps 14;
QY 11 TLRTSRSLRTPYSRDLRPPFTETATILEDPLPEEC-----NLTMHN 58
Db 24 TLHPVLELAARETL---RLSPEDI---VLYRVHVEEIIERSEFTETWNRFTHT 76
QY 59 VS-----YVRGLPCSVGFTLQEWVVPMDVLTREELVILRKCMHVC--LCCA 104
Db 77 ERVLDLNSVLEIFHRGDP-SLGRAL--AWMA-W-----CMHACRTLCN 118
QY 105 N-----IDMTSMI---HGYESWALHCHSCSPGSLQCIAGGQVLASWFRWVDGAMFN 155
Db 119 OSTPYVVDLSVRGMLSEGLDGV-IRHQ-----GG-----WSTLIENIPGS 161
QY 156 QRFYREVYNNMPKEYMFMSSVPMRGRH 185
Db 162 RRFSTWFLAGLTSLVICSYLFISGRH 191

RESULT 5

W48269 ID W48269 standard; Protein; 458 AA.
AC W48269;
DT 23-JUN-1998 (first entry)
DE Open reading frame RHKI protein SEQ ID NO:6 from J10075789 Example 2.
KW Killer protein; protease resistant; HM-1; RHKI; Hansenula mrakii;
KW yeast; mutant; antibacterial; soap.
OS Unidentified.
PN J10075789-A.
PD 24-MAR-1998.
PF 04-SEP-1996; 234393.
PR 04-SEP-1996; JP-234393.
PA (EIJT-) EIJIN KENKYUSHO KK.
DR WPI: 98-244362/22.
DR N-PSDB: V20655.
PT Hansenula mrakii protease resistant killer protein - has antibacterial properties and may be used as active ingredient in e.g. soap
PS Example 2; Page 10-12; 16pp; Japanese.
CC The present sequence represents RHKI used from an example of the present invention which describes a Hansenula mrakii protease resistant killer protein. The killer protein has antibacterial activity and so can be used in an antibacterial agent which contains the killer protein as the active component. The killer protein can also be used as the active component in an antibacterial soap. The protein, which was derived from the yeast Hansenula mrakii, has antibacterial properties and may be used in sterilising and cleaning compositions e.g. soap.
CC The protein has a high protease resistance and can be prepared recombinantly in a large amount.
SQ Sequence 458 AA;

Query Match 4.78; Score 75.5; DB 1; Length 458;
Best Local Similarity 19.58; Pred. No. 4.9;
Matches 31; Conservative 31; Mismatches 60; Indels 37; Gaps 9;

QY 59 VSYVRGLP-----CSVGF--TLIQEWVVPMDVLTREELVILRKCMHVCCLCANIDIMTS 111
Db 261 VPFLRSFPQQLHCAFNFGKFMQWSINQMM---DEAFNDKRFHLALISHLIALT 317
QY 112 MMINGY-----ESWALHCHSCSPGSLQCIAGGQ-----VLASWFRWVDGAMFNORF 158
Db 318 LFVTRYPRILPDLASLCH---PLRKNVLANPAKTIFFVLIASNFTGVLSRSLHYQF 374
QY 159 I-WYREVYNNMPKEYMFMSSVPMRGRHLIYLRWDGH 196
Db 375 LSWY-----HWTLPILIFWGMPPFVG-----PIWYVLH 403

RESULT 6

W01111 ID W01111 standard; Protein; 1255 AA.
AC W01111;
DT 01-JAN-1997 (first entry)
DE HER-2/neu protein.
KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase; breast cancer; ovary cancer; colon cancer; lung cancer; prostate cancer; immunisation; tumour; vaccine; vector.
OS Homo sapiens.
PN domain
PF 676..1255
PR Location/Qualifiers
PS /label= Intracellular_domain
PT /note= "claimed domain, useful for immunisation"
FN W09630514-A1.
PD 03-OCT-1996.
PF 28-MAR-1996; U01689.
PR 31-MAR-1995; US-414417.
PA (UNIW) UNIV WASHINGTON.
PI Cheever MA, Disis ML;
DR WPI: 96-455361/45.
DR N-PSDB: T40739.
PT DNA encoding HER-2-neu poly:peptide(s) - used for prevention or

PT treatment of malignancies with which the HER-2/neu oncogene is
 associated
 PS Claim 2: Page 56-61; 71pp; English.
 PT Human HER-2/neu protein (W0111), also called p185 or c-erbB2, is
 CC the product of the HER-2/neu oncogene (see also T40739). The
 CC protein is over-expressed in various cancers, including breast,
 CC ovarian, colon, lung and prostate. The intracellular domain of the
 CC protein can be used to immunise an animal against a malignancy with
 CC which the oncogene is associated. The polypeptide can be produced
 CC in transformed host cells for use in immunisation. Alternatively,
 CC animal cells are transfected in vivo or ex vivo with a viral vector
 CC that directs expression of the polypeptide.
 SQ Sequence 1255 AA;

[illegible]

RESULT	7
W92D46	
ID	W92406 standard; Protein; 1255 AA.
AC	W92406;
DT	21-APR-1999 (first entry)
DE	Human HER-2/neu oncogene protein.
KW	HER-2/neu; oncogene; immune response; T cell; B cell; immunisation; malignancy; treatment; tumour.
OS	Homo sapiens.
FH	Key
FT	Region
ET	Location/Qualifiers 676..1255
PN	/note= "region which elicits immune response"
PD	US5869445-A.
PP	09-FEB-1999.
PF	01-APR-1996; 625101.
PR	01-APR-1996; US-625101.
PR	17-MAR-1993; US-033644.
PR	12-AUG-1993; US-106112.
PR	31-MAR-1995; US-414417.
PA	(UNIW) UNIV WASHINGTON.
PI	Cheever MA, Disis ML;
DR	WPI: 99-152835/13.
DR	N-PDSB; X01912.
DPT	Use of HER-2/neu polypeptides - for eliciting an immune response to an HER-2/neu associated malignancy, particularly for treating or preventing tumours
PT	Claim 3; Column 31-38; 26pp; English.
PS	This sequence represents the human HER-2/neu oncogene protein. A fragment of this protein is used in a method for eliciting or enhancing an immune response.

CC response to HER-2/neu protein. The polypeptide can stimulate T cells and
CC B cells to produce an immune response to the HER-2/neu protein. The
CC method can be used for immunisation against a malignancy in which the
CC HER-2/neu oncogene is associated and in the treatment of an existing
CC tumour, or to prevent tumour occurrence or reoccurrence.
SQ Sequence 1255 AA;

[illegible]

RESULT	8	
W03549		
AC	W03549 standard; Protein; 265 AA.	
DT	W03549;	
DE	21-APR-1997 (first entry)	
DE	ORF-5 protein sequence from BamHI fragment of HVT.	
KW	Herpes virus of turkey; open reading frame; ORF; homology; vector;	
KW	avian herpes virus; recombinant viral vaccine; intergenic region; IBDV;	
KW	cytomegalovirus immediate early promoter; UL55 gene; repeat region; ILTV;	
KW	antigen; infectious bursal disease virus; Marek's disease virus; MDV;	
KW	infectious laryngotracheitis virus; avian anaemia virus; vaccination;	
KW	infectious bronchitis virus; IBV; poultry; Gumboro disease;	
KW	Newcastle disease.	
OS	Herpesvirus of turkey.	
PN	EP-719864-A2.	
PD	03-JUL-1996.	
PF	28-DEC-1995; 402970.	
PR	30-DEC-1994; FR-016017.	
PA	(INMR) RHONE MERIEUX SA.	
PI	Audonnet JCF, Bublout MM, Dartell RJ, Duinat CV;	
PI	Laplace ELF, Riviere MAE;	
DR	WPI; 96-364150/37.	
DR	N-PSDB; T39309.	
PT	Live recombinant avian vaccine - comprises herpes virus as vector	
PT	and having sequence encoding antigenic polypeptide inserted between	
PT	UL55 gene and repeat region	
PT	Example 4; Fig 1; 50pp; French.	
PS	This is the amino acid sequence of the protein encoded by the open	
CC	reading frame 5 (ORF-5) from the BamHI I fragment from the herpes virus	
CC	of turkeys (HVT). The DNA fragment was isolated from HVT strain FC126	
CC	(J. Gen. Virol., 70:1789-1804 (1989)) and is used as a basis for	
CC	generating live recombinant viral vaccines using the herpes virus DNA	
CC	sequences as vectors. The fragment contains 3 intergenic regions into	
CC	which genes encoding heterologous proteins to be expressed can be	
CC	inserted, preferably under control of the cytomegalovirus immediate early	
CC	(CMV-IE) promoter. The heterologous proteins can be used to vaccinate	

CC poultry against Gumboro disease (caused by IBDV), Newcastle disease,
 CC Marek's disease, infectious bronchitis, infectious laryngotracheitis and
 CC avian anemia.
 SQ Sequence 265 AA;

Query Match 4.6%; Score 73.5; DB 1; Length 265;
 Best Local Similarity 29.4%; Pred. No. 3.8;
 Matches 20; Conservative 9; Mismatches 28; Indels 11; Gaps 2;
 QY 89 ELVILAK--CMHYCLCANIDITSMHIGYEWALHCHCSSPGSLQCIAGGOVLASWFR 146
 DB 173 QLMDSKILCLISYACDN-----PGAQGMILSCYGRPGGLQCLAFIRTLMLFN 223
 QY 147 MVVDGAMF 154
 DB 224 DVLSSREF 231

RESULT 9
 W57324
 ID W57324 standard; Protein; 391 AA.
 AC W57324;
 DT 14-SEP-1998 (first entry)
 DE Cytosolic glycerol-3-phosphate dehydrogenase GPD1.
 KW Glycerol-3-phosphate dehydrogenase; G3PDH; GPD1; DAR1; OSG1;
 KW D2830; YDL022W; yeast.
 OS Saccharomyces sp.
 PN WO9821340-AL.
 PD 22-MAY-1998.
 PF 10-NOV-1997; U20293.
 PR 13-NOV-1996; US-030602.
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 PA (GEMV) GENENCOR INT INC.
 PI Bulthuis BA, Gatenby AA, Haynie SL, Hsu AK, Lareau RD;
 DR WPI: 98-297943/26.
 DT N-PSDB; V30822.
 PT Fermentative production of glycerol using recombinant host -
 containing genes for glycerol-3-phosphate dehydrogenase and/or
 glycerol-3-phosphatase
 PS Claim 9; Page 31-32; 5pp; English.
 CC This Saccharomyces polypeptide comprises a cytosolic
 CC glycerol-3-phosphate dehydrogenase (G3PDH) that catalyses the
 CC conversion of dihydroxyacetone phosphate to glycerol-3-phosphate.
 CC It is encoded by the GPD1 (DAR1, OSG1, D2830, YDL022W) gene (see
 CC V30822). The invention provides recombinant organisms that
 CC express G3PDH and/or glycerol-3-phosphatase (G3P) (see also
 CC W57325-32) useful for the production of glycerol from a variety of
 CC C-sources. A host cell is preferably transformed with a cassette
 CC containing either a G3PDH gene and/or a G3P gene and then cultured
 CC in the presence of a mono-, oligo-, polysaccharide or LC-substrate.
 CC The glycerol obtained is used in cosmetics, liquid soaps,
 CC pharmaceuticals, lubricants and antifreezes; its esters are used in
 CC the oil and fat industries. The method produces glycerol rapidly
 CC and inexpensively without generation of polluting by-products.
 SQ Sequence 391 AA;

Query Match 4.6%; Score 73; DB 1; Length 391;
 Best Local Similarity 23.2%; Pred. No. 7.4;
 Matches 45; Conservative 21; Mismatches 58; Indels 70; Gaps 11;

QY 45 HPLLPE-CNTLTMHNVSVIRGLPCSVGFTLIQWVVPDWMVLTREELVILRKMHVCLCC 103
 DB 127 HQFLPRICSQLKHVDHSHVRAISCLKGFEVGAQVQLSSVIT-EELGI-----QCGAL 179
 QY 104 ANIDITSMHIGYEWALHCHCSSPGSLQCIAGGOVLASWFRMVVDGAMFNQFIWYRE 163
 DB 180 SGANIATEV---AQEHWIS-----ET 196
 QY 164 VVNYNPKVEMFMSSVFMGR-----HLIYLRWYDG---HVGSVVPAMSGYSAHLCGIL 216
 DB 197 TVAYHIPKD-----FRGEGKVDHVKALKALFHRPYFHV-SVIEDVA-GISI--CGAL 244

QY 217 NNIVLCCSYCADL 230
 DB 245 KNVVALCGGFVEGL 258

RESULT 10
 W60255
 ID W60255 standard; Protein; 391 AA.
 AC W60255;
 DT 28-SEP-1998 (first entry)
 DE Klebsiella pneumoniae glycerol-3-phosphate dehydrogenase.
 KW Glycerol-3-phosphate dehydrogenase; production;
 KW 1,3-propanediol; recombinant.
 OS Klebsiella pneumoniae.
 PN WO9821341-A2.
 PD 22-MAY-1998.
 PF 13-NOV-1997; U20873.
 PR 13-NOV-1996; US-030601.
 PA (GEMV) GENENCOR INT INC.
 PI Chase MW, Diaz-torres M, Dunn-coleman NS, Trimbur D;
 DR WPI: 98-297944/26.
 DT N-PSDB; V35737.
 PT New method for increasing production of 1,3-propanediol - comprises
 PT fermentation of inexpensive carbon sources by microorganism
 PT expressing dehydratase, used, e.g. to prolong half-life of enzyme
 PS Disclosure; Page 65-66; 13pp; English.
 CC The sequence is that of cytosolic glycerol-3-phosphate dehydrogenase.
 CC It was used as part of a method of fermentative production
 CC of 1,3-propanediol (1,3-pd), using an organism comprising
 CC at least 1 gene encoding a dehydratase, is improved by
 CC inserting into the host a gene encoding protein X and culturing
 CC the transformant in presence of a carbon source (e.g. mono-, oligo-
 CC or poly-saccharide or LC substrate) convertible to 1,3-pd.
 CC 1,3-pd is a starting material for polyesters, polyurethanes and
 CC cyclic compounds. 1,3-pd can now be produced by a single
 CC recombinant organism from inexpensive carbon sources such
 CC as glucose (rather than costly glycerol or dihydroxyacetone),
 CC rapidly and without causing pollution.
 SQ Sequence 391 AA;

Query Match 4.6%; Score 73; DB 1; Length 391;
 Best Local Similarity 23.2%; Pred. No. 7.4;
 Matches 45; Conservative 21; Mismatches 58; Indels 70; Gaps 11;

QY 45 HPLLPE-CNTLTMHNVSVIRGLPCSVGFTLIQWVVPDWMVLTREELVILRKMHVCLCC 103
 DB 127 HQFLPRICSQLKHVDHSHVRAISCLKGFEVGAQVQLSSVIT-EELGI-----QCGAL 179
 QY 104 ANIDITSMHIGYEWALHCHCSSPGSLQCIAGGOVLASWFRMVVDGAMFNQFIWYRE 163
 DB 180 SGANIATEV---AQEHWIS-----ET 196
 QY 164 VVNYNPKVEMFMSSVFMGR-----HLIYLRWYDG---HVGSVVPAMSGYSAHLCGIL 216
 DB 197 TVAYHIPKD-----FRGEGKVDHVKALKALFHRPYFHV-SVIEDVA-GISI--CGAL 244
 QY 217 NNIVLCCSYCADL 230
 DB 245 KNVVALCGGFVEGL 258

RESULT 11
 W30680
 ID W30680 standard; Protein; 391 AA.
 AC W30680;
 DT 12-OCT-1998 (first entry)
 DE Glycerol-3-phosphate dehydrogenase GPD1.
 KW Glycerol-3-phosphate dehydrogenase; GPD1; 1,3-propanediol.
 OS Saccharomyces sp.
 PN WO9821339-A1.
 PD 22-MAY-1998.

Mon Jun 26 09:10:34 2000

10-NOV-1997; U20292.
 13-NOV-1996; US-030601.
 (DUPO) DU PONT DE NEMOURS & CO E I.
 (GEMV) GENECOR INT INC.
 Dias-Torres M, Gatenby AA, Haynie SL, Hsu AK, Lareau RD,
 Nagarajan V, Nair RV, Nakamura CE, Payne MS, Picataggio SK,
 Trimbur DE, Whited GM;
 WPI; 98-297942/26.
 N-PSDB; V42016.
 Fermentative production of 1,3-propanediol - by single organism
 containing cassette comprising specific genes, and capable of using
 inexpensive carbon sources
 claim 9; Page 55-56; 95pp; English.
 Cytosolic glycerol-3-phosphate dehydrogenase Gppl is encoded by the
 gppl gene (see V42016) of Saccharomyces sp. This enzyme catalyses
 the conversion of dihydroxyacetone phosphate to glycerol-3-phosphate.
 A claimed method for production of 1,3-propanediol (1) comprises
 culturing a microorganism transformed with a cassette containing at
 least 1 of the genes (see V42012-21) for glycerol-3-phosphate and
 dehydrogenase, glycerol-3-phosphatase, glycerol dehydratase and
 1,3-propanediol oxidoreductase (see W30676-85). Any of these genes
 not present on the cassette must be present endogenously. The
 enzyme sequences may include substitutions, deletions and additions
 provided activity is not altered. A single recombinant organism
 can now be used for production of (1) from inexpensive C-sources
 CC (contrast use of glycerol or dihydroxyacetone) without causing
 CC pollution. (1) is a starting material for polyesters, polyurethanes
 CC and cyclic compounds.
 CC Sequence 391 AA;
 SO
 Query Match:
 Best Local Similarity 4.6%; Score 73; DB 1; Length 391;
 Matches 45; Conservative 21; Mismatches 58; Indels 70; Gaps 11;
 QY 45. HPLPE-CNTLTMHNSVYRGLPCSVGFTLQEVWVMDVLTREELVILKMKHVCVLC 103
 DB 127 HQFLPRICSLKGVHDSHVRAISLKGFEVGAQVQLSSYIT-DELGI-----CCGAL 179
 QY 104 ANTIDTSMHGHYESWALHCHCSSPGSLQCIAGQVLASWFRMVVDGAMFNQRFIYRE 163
 DB 180 SGANIATEV---AQEHW-----ET 196
 QY 164 VVYNNPKFEMFMSVFMGR-----HLIYRLWYDG--HVGSVVPAMSGYSALHCGIL 216
 DB 197 TVAYHIPK-----FRGSGKVDHVKLVKALFRPYFHV-SVIEDVA-GISI--CGAL 244
 QY 217 NNIVLCCSYCADL 230
 DB 245 KNYVALGCGFEVGL 258
 RESULT 12
 ID R08222 standard; protein; 624 AA.
 AC R08222;
 DT 06-MAR-1991 (first entry)
 DE Extracellular portion of the human epidermal growth factor receptor 2.
 KW Human epidermal growth factor receptor 2; HER2; vaccine; cancer.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT region 22..32
 FT /label= epitope
 FT /note= "potential T-cell epitope"
 FT region 74..84
 FT /label= epitope
 FT /note= "potential T-cell epitope"
 FT region 113..130
 FT /label= epitope
 FT /note= "potential T-cell epitope"
 FT region 187..197
 FT /label= epitope
 FT /note= "potential T-cell epitope"

FT region 305..315
 FT /label= epitope
 FT /note= "potential T-cell epitope"
 FT region 327..353
 FT /label= epitope
 FT /note= "potential T-cell epitope"
 FT region 375..385
 FT /label= epitope
 FT /note= "potential T-cell epitope"
 FT region 398..408
 FT /label= epitope
 FT /note= "potential T-cell epitope"
 FT region 433..443
 FT /label= epitope
 FT /note= "potential T-cell epitope"
 FT region 519..529
 FT /label= epitope
 FT /note= "potential T-cell epitope"
 FT WO9014357-A.
 PD 29-NOV-1990.
 PF 18-MAY-1990; U02697.
 PR 19-MAY-1989; US-354319.
 PA (GETH) GENENTECH INC.
 PI Hudziak RM, Shepard HM, Ullrich A;
 DR WPI; 90-375946/50.
 DR N-PSDB; Q06828.
 PT HER2 extracellular domain used as vaccine - comprises sequence of
 at least 9 amino acid(s) prep. using expression vector of DNA
 isolated from human epidermal growth factor receptor
 PS Disclosure; Fig 13; 49pp; English.
 CC This claimed sequence is free of transmembrane or intracellular
 CC portions of the HER2 mol. The protein is antigenic in animals.
 CC Fragment comprising at least 9 amino acids open several
 CC possibilities for further research and a broad range
 CC of potential therapeutic applications, e.g. for the treatment of
 CC mammary gland adenocarcinoma and other cancers.
 CC Sequence 624 AA;
 SO
 Query Match:
 Best Local Similarity 4.5%; Score 72; DB 1; Length 624;
 Matches 52; Conservative 24; Mismatches 92; Indels 96; Gaps 12;
 QY 20 SRRTPYSRDLRPPFTETRAT-----ILEDHPLPECNLT-----TMHNSV---VR 63
 DB 366 SNTAPLOPEQLQVFFLEITGTYLISAWRDSLPDLVSFQNLQVIRGRLHNGAYSLTLQ 425
 QY 64 GLPCS-----VGFTLIQE-----VVVPMVMVLTREELVILR-----KCMH 98
 DB 426 GLGISWGLRSLRELGSGLALHHTHLCFVHTVPWQDLFRNPHQALLHTANRPEDECVG 485
 QY 99 VCLCCANIDMTSMHGHYESWALHCHCSSPGSLQCIAGQVLASWFRMVVDGAMFNQRF 158
 DB 486 EGLACHQL-----CARGHCGPGTQCVNCSQ-----F 513
 QY 159 IWTREVVN-----YNNPKFEMFMSVFMGRHLIYRLWYDGHVGVSVVPAMSGYSALH 213
 DB 514 LRQCEVEECRVLQGLPRE-----YVNAHCLPCHPECPQNGSVT---CFGPEADQC 563
 QY 214 GILNNIVLCCSYCADLSEIRVRC 237
 DB 564 -----VACAHYKDPFPCVARC 579
 RESULT 13
 W62647
 ID W62647 standard; protein; 369 AA.
 AC W62647;
 DT 09-OCT-1998 (first entry)
 DE Mature durum wheat glutenin protein.
 KW Glutenin gene; durum wheat; low-molecular-weight;
 KW transgenic durum wheat.
 OS Triticum durum.


```

352  LTNNGIAGVCKIRPTEELATADEDFRRGYAEYFENKPKPLMHYSVISGFFGDHTKIPN 411
      QY      51  CNTLTW-HNVSVRGLPCSVGFTLIQE-----VVVPKDWLTREE 89
      Db      412  GFEMTFHELEY----PFSRGVRTISANPYDAPDFDGFNLDERDLMPWYAKKSRET 467
      QY      90  LVILRK-----MHVCLCCANIDIMTSMTHGYESWALHCHC 126
      Db      468  ARMESFAGEVTSHHPLFKVDSPAROLDLETCSAYAGPKHLTANTLYHG--SWTVPIDK 525
      QY      127  SSPGSLQIAGQV 140
      Db      526  PTPKNDHFVTSNOV 539

```

RESULT	15
R54843	
ID	R54843 standard; Protein; 1058 AA.
R54843	
AC	
DE	HER4 with alternate 3'-end without AP domain.
DT	11-JAN-1995 (first entry)
DE	Human epidermal growth factor receptor; HER; EGF; tyrosine kinase;
KW	cancer; neuronal tissue; muscle tissue; neoplasm;
KW	carcinoma; primer; probe; PCR; autophosphorylation domain; AP.
OS	Homo sapiens.

N-PSDB: Q64898.
New recombinant nucleic acid expressing HER4 - a new receptor tyrosine kinase expressed in some cancer cells, and related vectors, antibodies, ligands etc, for diagnosis and treatment of cancers
Claim 12; Fig 2A; 104pp: English.
The DNA encoding this sequence is identical with that of HER4 (Q64898) up to nucleotide 3168, where the sequence diverges and the ORF stops after 13 amino acids, followed by an extended, unique 3'-UTR.
HER4 is the fourth member of the EGFR-family of tyrosine kinases and is expressed in some human cancers and in some tissues of neuronal or muscle origin. HER4 polynucleotides, opt. labelled, are useful in assays (e.g. of HER4 mRNA to detect certain neoplasms, esp. breast carcinoma) and as primers in PCR or as probes.
Sequence 1058 AA:
SSQ

Search completed: June 23, 2000, 10:03:34
Job time: 3362 sec

us-09-214-478-2.rai

RESULT 3
US-08-467-083-68
; Sequence 68, Application US/08467083
; Patent No. 5726023
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Dials, Mary L.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/NEU PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH
; TITLE OF INVENTION: HER-2/NEU ONCOGENE IS ASSOCIATED
; TITLE OF INVENTION: HER-2/NEU ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESS:

STREET: 6300 Columbia Center, 701 14th Avenue
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,083
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/414,417
FILING DATE: 06-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Sharkey, Richard G.
REGISTRATION NUMBER: 32,629
REFERENCE/DOCKET NUMBER: 920010.448C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 1255 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-467-083-68

	Query Match	5.4%;	Score 86;	DB 1;	Length 1255;
	Best Local Similarity	18.5%;	Pred. No. 0.42;		
	Matches 65;	Conservative	34;	Mismatches 116;	Indels 136; Gaps
QY	20	SRRTPYSRDLRPFFETETRAT----	ILEDHPHLPECNLT-----	TMHNVS-----VR	63
		:: ::		:: ::	:
Dd	387	SNTAPLQPOLQVFETLEETGTYLISAWPDSLPDLSVFNQLVIRGRILHNGAYSLTIQ	446		
QY	64	GLPCS-----VGFTLIQE-----	WVPWDMLTRELVLIR-----	KCMH	98
		:: ::		:: ::	:
Dd	447	GLGISWLGRSLRELGSGALIHNNHTLCFVHTVPWDQLFERNPHQALLHTANRPEDCVG	506		
QY	99	VCLCCANDIMTSMIHHGESWALHCHCSSPGSLQCIAGGQVLASFWRMVVDGAMENQRF	158		
		:: ::		:: ::	-F
Dd	507	EGLACHQL-----CARCHCWGGPTQCNCQ-----			534
QY	159	IWYREVVN-----YNMKVEVMFMSSVEMRGRLHYLRWLMDGHGVSVVPAMSGYSALHC	213		
		:: ::		:: ::	
Dd	535	LRGQECVEBCRVTLGLPRE-----YVNARHCLPCHECQPNQSVT-----	CFGEADQC	584	
QY	214	-----GILNNIVLVCCSYCADLSEIRVRC	237		
		:: ::		:: ::	:
Dd	595	VACAHKDPFCVCARCPGVKPRDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDD--KG	641		
QY	238	CARRTR-----RLMLRAVIIAETAMLYSCRTERRRQQF-----	IRALLQ	279	
Dd	642	CPAFORASPLTISIISAVVGIILVLLVWGVSFGIILKRROOKIKRYTMRRLLQ	692		

RESULT 5
US-08-486-348A-68
; Sequence 68, Application US/08486348A
; Patent No. 5846538
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Disis, Mary L.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486.348A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-486-348A-68

Query Match 5.4%; Score 86; DB 2; Length 1255;
Best Local Similarity 18.5%; Pred. No. 0.42;
Matches 65; Conservative 34; Mismatches 116; Indels 136; Gaps 15;

QY 20 SRRTPYSRDLPPFETETRA-----ILEDHPLLPECNTL-----TMHNVSY---VR 63
Db 387 SNTAPLOPEQLQVFETLEETGYLYISAWPDSLPLSVFQNLQVIRGRILHNGAYSLTLQ 446
QY 64 GLPCS-----VGFTLQIE-----WVVPDMLVLTREELVILR-----KCMH 98
Db 447 GLGISWGLRLSRLGSLALIHHTHLCFVHTVPDQLFRPHQALLHTANRPEDECVG 506
QY 99 VCLCCANIDIMTSMHIGYESWALHCHSCSPGSLQCIAGGQVLAWSFRMVVDGAMFNORF 158
Db 507 EGLACHQL-----CARCHCWGPPTQCNCQ-----F 534
QY 159 IWTREYVN-----YNNPKYMFMSVFMGRHLIYLRWLDGHVGSVVPAMSGYSALHC 213
Db 535 LRQECVEECRVLQGLPRE-----YNNARHCLPCHPECPQNGSVT---CFGPEADQC 584
QY 214 -----GILNNIVVLCSCYCADLSEIRVRC 237
Db 585 VACAHYKDPFPCVARGPSGVKPDLSYMPINKFPDEGACQPCINCHSCVDLDD---KG 641
QY 238 CARTR-----RLMLRAVRIIAETTAMLYSCRTERROOF-----IRALLQ 279
Db 642 CPAEQRASPLTSIISAVVGLLVVGLVGVFGILIKRQOKIRKVTMRRLQ 692

RESULT 6
US-08-468-545B-68

RESULT 4
US-08-414-417B-68
; Sequence 68, Application US/08414417B
; Patent No. 5801005
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Disis, Mary L.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/414.417B
; FILING DATE: 31-MAR-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-414-417B-68

Query Match 5.4%; Score 86; DB 1; Length 1255;
Best Local Similarity 18.5%; Pred. No. 0.42;
Matches 65; Conservative 34; Mismatches 116; Indels 136; Gaps 15;

QY 20 SRRTPYSRDLPPFETETRA-----ILEDHPLLPECNTL-----TMHNVSY---VR 63
Db 387 SNTAPLOPEQLQVFETLEETGYLYISAWPDSLPLSVFQNLQVIRGRILHNGAYSLTLQ 446
QY 64 GLPCS-----VGFTLQIE-----WVVPDMLVLTREELVILR-----KCMH 98
Db 447 GLGISWGLRLSRLGSLALIHHTHLCFVHTVPDQLFRPHQALLHTANRPEDECVG 506
QY 99 VCLCCANIDIMTSMHIGYESWALHCHSCSPGSLQCIAGGQVLAWSFRMVVDGAMFNORF 158
Db 507 EGLACHQL-----CARCHCWGPPTQCNCQ-----F 534
QY 159 IWTREYVN-----YNNPKYMFMSVFMGRHLIYLRWLDGHVGSVVPAMSGYSALHC 213
Db 535 LRQECVEECRVLQGLPRE-----YNNARHCLPCHPECPQNGSVT---CFGPEADQC 584
QY 214 -----GILNNIVVLCSCYCADLSEIRVRC 237
Db 585 VACAHYKDPFPCVARGPSGVKPDLSYMPINKFPDEGACQPCINCHSCVDLDD---KG 641
QY 238 CARTR-----RLMLRAVRIIAETTAMLYSCRTERROOF-----IRALLQ 279
Db 642 CPAEQRASPLTSIISAVVGLLVVGLVGVFGILIKRQOKIRKVTMRRLQ 692

RESULT 7
US-08-625-101-2
; sequence 2, Application US/08625101
; Patent No. 5869445

```

Query Match          4.6%; Score 74; DB 2; Length 1255;
Best Local Similarity 18.2%; Pred. No. 11;
Matches 64; Conservative 34; Mismatches 117; Indels 136; Gaps 15;

QY 20 SRRTPYSRDLRPPYETETRAT-----ILEDHPLLPECNTL-----TMENVSY---VR 63
DB 387 SNTAPLOPEQLQVFETLEEITGYIISAWPDSLFDLSVFNQLQVIRGRILLNGAYSLTLO 446
QY 64 GLPCS-----VGFTLIQE-----WVVPDMVLPRELVILR-----KCMH 98
DB 447 GLGISWLGLRLRELGSGLALIHNTLCFVHTVPWDQLFRNPHQALLHTANRPEDECVG 506
QY 99 VCLCCANIDIMTSMKHGYESWALHCHSCSPGSLQCTAGGOVLASWFRMVVDGAMFNQRF 158
DB 507 EGLACHQL-----CARGCHWGEPGTCVNCNQ-----F 534
QY 159 IWRREVNW-----YNMPKEVMFMSVFNMRGRHLIYLRMYDGHVGSVVPAMSGFYSALHC 213
DB 535 LRQGEVVEECRVLQGLPRE-----YVNARHCLPCHPECPQNGSVT---CFGPEADQC 584
QY 214 -----GILNNIVLCCSYCADLSIEVRVC 237
DB 585 VACAHYKDPPEFCVARGSPGVKPDLSYMPDWKFPDEEGACQPCPNCTHSCVDLDD---KG 641
QY 238 CARRTR-----RLMLRAVRIIAETETALLYSCRTERRRQQF-----IRALLQ 279
DB 642 CPAQRASPLSIISAVVGILLVVVLGVVFGILIKRRQOKIKYTMRLILO 692

RESULT 8
US-08-356-786-2
: Sequence 2, Application US/08356786
: Patent No. 5877305

```

GENERAL INFORMATION:

APPLICANT: Huston, James S.
APPLICANT: Oppermann, Hermann
APPLICANT: Houston, L. L.
APPLICANT: Ring, David B.
TITLE OF INVENTION: Biosynthetic Binding Protein for Cancer
NUMBER OF INVENTION: Marker
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Edmund R. Pitcher, Testa, Hurwitz, & Thibault
STREET: Exchange Place, 53 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/356,786

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/831,967

FILING DATE: 06-FEB-1992

ATTORNEY/AGENT INFORMATION:

NAME: Pitcher, Edmund R.

REGISTRATION NUMBER: 27,829

REFERENCE/DOCKET NUMBER: CRP-053

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 248-7000

TELEFAX: (617) 248-7100

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1255 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-356-786-2

Query Match 4.6%; Score 74; DB 2; Length 1255;
Best Local Similarity 18.2%; Pred. No. 11;
Matches 64; Conservative 34; Mismatches 117; Indels 136; Gaps 15;

QY 20 SRTPTSRDLPPFETETRAT-----ILEDHPLLPQENTL-----TMHNVSY---VR 63
DB 387 SNTAPLQPEQLQVFTLEETITGLYXISAWPDSLPLSVFQNLQVIRGRILHNGAYSILQ 446
QY 64 GLPCS-----VGTLLQ-----WVVPDMVLTREELVILR-----KCMH 98
DB 447 GLGISWGLRSRELGSLALHHTHLCFVITVPDQLFRNPHQALLHTANRPEDECVG 506
QY 99 VCLCANTIDMTSMIHGYESWALHCHCSSPGSLQIAGGQVLASWFRVVDGAMENQRF 158
DB 507 EGLACHQL-----CARGHCWGPQTQVNCQ-----F 534
QY 159 IYREVVN-----YNNKPEVMEFSSVFRGRLIYLRLWDGHVGSVVPAMSGYSALHC 213
DB 535 LRGCVEECRVLOGLPRE-----YVNARHCLPCHPECPQNGSVT---CFGPEADQC 584
QY 214 -----GILNNIVLCCSYCADLSEIRVRC 237
DB 585 VACAHKDPFCVAPCSVKPDLSPMTWPKFPDEGACQCPINCTHSCVDLDD---KG 641
QY 238 CARTR-----RLMLRAVRIIAETMTAMLYSCRTERRRQF-----IRALLQ 279
DB 642 CPASORASPLTSIISAVVGLLVVGLVGVFGLIKRRQOKIRKYTMRLQL 692

RESULT 9

US-08-968-563-11
Sequence 11, Application US/08968563
Patent No. 6013494
GENERAL INFORMATION:
APPLICANT: CHARLES E. NAKAMURA
APPLICANT: ANTHONY A. GATENBY
APPLICANT: AMY (KUANG-HUA) HSU
APPLICANT: RICHARD D. LA REAU
APPLICANT: SHARON L. HAYNIE
APPLICANT: MARIA DIAZ-TORRES
APPLICANT: DONALD E. TRIMBUR
APPLICANT: GREGORY M. WHITED
APPLICANT: VASANTHA NAGARAJAN
APPLICANT: MARK S. PAYNE
APPLICANT: STEPHEN K. PICATAGGIO
APPLICANT: RAMESCH V. NAIR
TITLE OF INVENTION: METHOD FOR THE RECOMBINANT
TITLE OF INVENTION: PRODUCTION OF 1,3-PROPANEDIOL
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
ADDRESSEE: GENENCOR INTERNATIONAL, INC.
STREET: 4 CAMBRIDGE PLACE
STREET: 1870 SOUTH WINTON ROAD
CITY: ROCHESTER
STATE: NEW YORK
COUNTRY: U.S.A.
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 50 INCH DISKETTE
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,563
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/030,601
FILING DATE: NOVEMBER 13, 1996
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CR-9982
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: GPDI
US-08-968-563-11

Query Match 4.6%; Score 73; DB 3; Length 391;
Best Local Similarity 23.2%; Pred. No. 2.5;
Matches 45; Conservative 21; Mismatches 58; Indels 70; Gaps 11;

QY 45 HPLLE-CNTLTMHNVSYVRGLPCSVGFLIOEWVVDKMLTRELIVLRKCMHVCCLC 103
DB 127 HFLPRICSQLKGHVDHSHVRAISCLGFEVGARGVOLLSSYIT-EELGI-----QCGAL 179
QY 104 ANIDIMTSMIHGYESWALHCHCSSPGSLQIAGGQVLASWFRVVDGAMENQRFIYRE 163

```

DD- 180 SCANITATEV---AQSHWS-----ET 196
OY- 164 VVNYNMPVFMFMSSVFMRGR-----HLYLRLWYDG---HVGSVVPAMSFYSALHCGIL 216
DD- 197 TVAYHPIKD-----FPEGCKVDHVKLALFHPFYFHV-SVIEDVA-GIST--CGAL 244
OY- 217 NNIVLCCSYCADL 230
DB- 245 KNVVALGCGFEVGL 258

```

RESULT 10
 US-08-991-300-2
 ; Sequence 2, Application US/08991300
 ; Patent No. 5973225
 ; GENERAL INFORMATION:
 ; APPLICANT: D'OVIDIO, RENATO
 ; APPLICANT: PORCEDDU, ENRICO
 ; APPLICANT: MERCHITELLI, CINZIA
 ; APPLICANT: CARDELLI, LUISA ERCOLI
 ; TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF A GENE
 ; TITLE OF INVENTION: ENCODING A LOW MOLECULAR WEIGHT GLUTENIN
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
 ; ADDRESSEE: P.C.
 ; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY
 ; CITY: ARLINGTON
 ; STATE: VA
 ; COUNTRY: USA
 ; ZIP: 22202
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/991,300
 ; FILING DATE: 16-DEC-1997
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: IT MI 96/A 002663
 ; FILING DATE: 19-DEC-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: OBLON, NORMAN F.
 ; REGISTRATION NUMBER: 24,618
 ; REFERENCE/DOCKET NUMBER: 2264-0201-0X
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 703-413-3000
 ; TELEFAX: 703-413-2220
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 369 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-991-300-2

```

Query Match          4.5%; Score 71; DB 2; Length 369;
Best Local Similarity 21.3%; Pred. No. 4;
Matches 29; Conservative 18; Mismatches 49; Indels 40; Gaps 7;

QY      7  PFGMTLRPRSR---LSRRTPYSRDLRPPETETRAILEDHLLP----- 50
      ||      |      |      |      |      |      |      |      |
Db      119 PFSQQLPPFSQQQPVLPQPFQSQQPPFSQQLPPFSQQQPVLPQOQPFSSQQQPIP 178
QY      50 -----ECNLTLMHNVSVRGL-----PCSVGLTIQEWVWDM--VLTREELVI 92
      ||      |      |      |      |      |      |      |      |
Db      179 PQQPPFSQQQQPVLLQQQIFVHPHSIIQLQNPCKV---FLQQQSPWMPQSLARSQMLQ 235
QY      93 LRKCMHYC--LCCANI 106

```

[illegible]

APPLICATION NUMBER: US/08/484,438
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/323,442
FILING DATE: 14-OCT-1994
APPLICATION NUMBER: US 08/150,704
FILING DATE: 10-NOV-1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/981,165
FILING DATE: 24-NOV-1992
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 5624-230
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1058 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-438-4

Query Match 4.4%; Score 70.5; DB 2; Length 1058;
Best Local Similarity 24.5%; Pred. No. 21;
Matches 27; Conservative 20; Mismatches 46; Indels 17; Gaps 4;

QY 69 VGFTLQEWVPMVLTREELVILRKCMH-----VCLCCANIDIMTSMHIGYESW 120
DB 748 VAIKILNETTGPKANVEFMDALIMASMDHPLVRLVGCL-SPTIOLVTLMPHGCLLE 806
QY 121 ALHCHCSSPGSLQCIAGGQVLASWFRMVVDGAMF-NQRFITWYREVYNNM 169
DB 807 YVHEKDNIGS-----QLLNWCQVIAKGMYLEERLVRDLAARNV 849

RESULT 13
US-08-484-438-2
Sequence 2, Application US/08/484438
Patent No. 5811098 5780031
GENERAL INFORMATION:
APPLICANT: Plowman, Gregory D.
APPLICANT: Culouscou, Jean-Michel
APPLICANT: Shoyab, Mohammed
APPLICANT: Siegall, Clay B.
APPLICANT: Hellstr m, Ingegerd
TITLE OF INVENTION: HER4 HUMAN RECEPTOR TYROSINE KINASE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,438
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/323,442
FILING DATE: 14-OCT-1994
APPLICATION NUMBER: US 08/150,704
FILING DATE: 10-NOV-1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/981,165
FILING DATE: 24-NOV-1992
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 5624-230
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1308 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-438-2

Query Match 4.4%; Score 70.5; DB 2; Length 1308;
Best Local Similarity 24.5%; Pred. No. 29;
Matches 27; Conservative 20; Mismatches 46; Indels 17; Gaps 4;

QY 69 VGFTLQEWVPMVLTREELVILRKCMH-----VCLCCANIDIMTSMHIGYESW 120
DB 748 VAIKILNETTGPKANVEFMDALIMASMDHPLVRLVGCL-SPTIOLVTLMPHGCLLE 806
QY 121 ALHCHCSSPGSLQCIAGGQVLASWFRMVVDGAMF-NQRFITWYREVYNNM 169
DB 807 YVHEKDNIGS-----QLLNWCQVIAKGMYLEERLVRDLAARNV 849

RESULT 14
US-08-390-000A-8
Sequence 8, Application US/08390000A
Patent No. 5985583
GENERAL INFORMATION:
APPLICANT: Sealfon, Stuart C.
TITLE OF INVENTION: Cloning and Expression of
TITLE OF INVENTION: Gonadotropin-Releasing Hormone Receptor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/390,000A
FILING DATE: 17-FEB-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6923-052
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-390-000A-8

Query Match 4.4%; Score 70; DB 2; Length 348;

Best Local Similarity 20.9%; Pred. No. 4.8;
Matches 23; Conservative 20; Mismatches 37; Indels 30; Gaps 4;

QY 81 WDMVLTREELVILKCMHVCLCANIDIMTSMIHYE-SWALHCHCSPGSLQCIAGGQ 139
DB 126 WSLVLAIERVYV-----VCKPMSNFRFGENHAIMGVAFVWVALACAP----- 171
QY 140 VLASFWRMVVDG-----AMFNORFIWYREVYNNMPKEYMF 175
DB 171 PLAGWSRYIPEGLQCSGIDYITLKPEVNNESFVIYMFVYHFTIPMIIF 220

RESULT 15

US-08-422-108-1
Sequence 1, Application US/08422108
Patent No. 6015567

GENERAL INFORMATION:

APPLICANT: Hudziak, Robert M.
APPLICANT: Shepard, H. Michael
APPLICANT: Ullrich, Axel
TITLE OF INVENTION: HER2 EXTRACELLULAR DOMAIN
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/422,108
FILING DATE: 14-Apr-1995
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/355460
FILING DATE: 13-DEC-1994
APPLICATION NUMBER: 08/048346

FILING DATE: 15-APR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/354319
FILING DATE: 19-MAY-1989
ATTORNEY/AGENT INFORMATION:

NAME: Lee, Wendy M

REGISTRATION NUMBER: 00,000

REFERENCE/DOCKET NUMBER: 554C2D2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/225-1994

TELEFAX: 415/952-9881

TELEX: 910/371-7168

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 624 amino acids

TYPE: Amino Acid

TOPOLOGY: Linear

US-08-422-108-1

Query Match 4.4%; Score 70; DB 3; Length 624;
Best Local Similarity 19.7%; Pred. No. 11;
Matches 52; Conservative 24; Mismatches 92; Indels 96; Gaps

QY 20 SRRTPYSRDRLPFFETETRAT-----ILEDHPLPECNLT-----TMHNVSY---VR 63
DB 366 SNTAPLQPEQLQVFETLEEITEYLYISAWPDSLPDLVSFQNLQVIRGRILHNGAISLILQ 425
QY 64 GLPCS-----VGFTLIQE-----WVVPDMVLTREELVILR-----KCMH 98
DB 426 GLGISWLGRLSRLGSLALIHHTHLCFVHVVPDQLFRNPHQALLHTANRPEDECVG 485
QY 99 VCLCCANIDIMTSMIHYE-SWALHCHCSPGSLQCIAGGQVLASFWRMVVDGAMFNORF 158
DB 486 EGLACHQL-----CARGHCWGPPTQCVCNSQ-----F 513
QY 159 IWYREVNV-----YNNPKEVFMSSVEMRGRHLIYLRWDYDGHVGSVVPAMSFYGSALHC 213
DB 514 LRGOECVEECRVLOGLPRE-----YVNARHCLPCHPECPQNGSVT---CFGPEADQC 563
QY 214 GILNNIVVLCSSYCADLSEIRVRC 237
DB 564 -----VACAHYKDPFPCVARC 579

Search completed: June 23, 2000, 10:04:33
Job time: 2838 sec

Mon Jun 26 09:10:35 2000

us-09-214-478-2.rai

Page 9

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2000, 09:36:35 ; Search time 61.02 seconds

(without alignments)
282.484 Million cell updates/sec

Title: US-09-214-478-2

Perfect score: 1594

Sequence: 1 MTTSGVPFGMTLRPTRSRLS.....RALLQHRPILMHYDSTPM 294

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 168808 seqs, 58629743 residues

Total number of hits satisfying chosen parameters: 168808

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

PIR_63:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1594	100.0	294	1 Q4ADC2	early E4 34K prote
2	858.5	53.9	291	2 S10867	early E4 34K prote
3	331	20.8	153	1 Q4ADB2	early E4 17K prote
4	186.5	11.7	288	1 Q4ADM1	early E4 33K prote
5	88.5	5.6	491	1 Q0B84	BHRF1 protein - hu
6	87	5.5	494	2 T32644	hypothetical prote
7	85.5	5.4	211	1 A70390	formate dehydrogen
8	82	5.1	1603	2 T24098	hypothetical prote
9	81.5	5.1	354	1 T51200	rhodopsin - Africa
10	79.5	5.0	456	2 T16796	hypothetical prote
11	79.5	5.0	562	2 T26242	hypothetical prote
12	79	5.0	352	1 S40688	opsin, rod - Pomat
13	79	5.0	459	2 B58851	NADH dehydrogenase
14	78	4.9	378	2 S36506	E2 protein - human
15	77.5	4.9	459	2 T11502	NADH dehydrogenase
16	77.5	4.9	530	2 T30305	hypothetical prote
17	77.5	4.9	642	2 T35003	probable respirato
18	77	4.8	374	2 T05240	pectate lyase (EC
19	77	4.8	446	2 S01187	NADH dehydrogenase
20	76.5	4.8	354	1 A48191	opsin, ultraviolet
21	76.5	4.8	459	1 QXBOAM	NADH dehydrogenase
22	76.5	4.8	459	2 T11411	NADH dehydrogenase
23	76	4.8	505	2 A35993	GTPase (EC 3.1.5.
24	76	4.8	528	2 T05131	hypothetical prote
25	75.5	4.7	373	2 T24487	hypothetical prote
26	75.5	4.7	458	2 S45424	ALG3 protein - yea
27	74.5	4.7	303	2 T32277	hypothetical prote
28	74.5	4.7	321	2 S65225	probable membrane
29	74.5	4.7	895	2 T23191	hypothetical prote
30	74	4.6	196	2 A40623	heat shock protein

31 74 4.6 1255 1 A24571 protein-tyrosine k
32 73.5 4.6 346 2 T14056 NADH dehydrogenase
33 73.5 4.6 446 2 I25797 NADH dehydrogenase
34 73.5 4.6 475 2 C64757 probable iron-sulf
35 73.5 4.6 604 1 S00726 protein kinase A-r
36 73.5 4.6 606 1 TVHUAF hypothetical prote
37 73 4.6 367 2 T25821 glycerol-3-phospha
38 73 4.6 391 2 S40059 glucosylceramidase
39 73 4.6 515 1 A32931 complement C3a ana
40 72.5 4.5 352 1 S27357 NADH dehydrogenase
41 72.5 4.5 459 2 S41829 NADH dehydrogenase
42 72.5 4.5 459 2 T11059 NADH dehydrogenase
43 72.5 4.5 459 2 T11149 ABC transporter sl
44 72.5 4.5 790 1 S77032 phospholipase C ho
45 72.5 4.5 1898 2 T42440

ALIGNMENTS

RESULT 1
Q4ADC2
early E4 34K protein - human adenovirus 2
C:Species: Mastadenovirus h2 (human adenovirus 2)
A:Note: Host Homo sapiens (man)
C:Date: 02-Apr-1982 #sequence_revision 02-Apr-1982 #text_change 04-Mar-1994
C:Accession: A03805
R:Herisse, J.; Rigolet, M.; Dupont de Dinechin, S.; Galibert, F.
Nucleic Acids Res. 9, 4023-4042, 1981
A:Title: Nucleotide sequence of adenovirus 2 DNA fragment encoding for the carboxylic
A:Reference number: A93733; M0ID:82059444
A:Accession: A03805
A:Molecule type: DNA
A:Residues: 1-294 <HER>
A:Note: this probable protein was assigned by correlating EM data and S1 digestion st
C:Genetics:
A:Map position: 92.6-95.2
C:Superfamily: adenovirus early E4 34K protein
C:Keywords: early protein

Query Match 100.0%; Score 1594; DB 1; Length 294;
Best Local Similarity 100.0%; Pred. No. 9.3e-142;
Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTTSGVPFGMTLRPTRSRLSRRTPYSRDLPPETETETRAILEDHPLPCCNTILTMHNS 60

Db 1 MTTSGVPFGMTLRPTRSRLSRRTPYSRDLPPETETETRAILEDHPLPCCNTILTMHNS 60

QY 61 YVRGLPCSVGFTLIQEVVVPMDVLTREELVILRKCMHVCLCCANIDIMTSMIHYESW 120

Db 61 YVRGLPCSVGFTLIQEVVVPMDVLTREELVILRKCMHVCLCCANIDIMTSMIHYESW 120

QY 121 ALHCHCSSPGSLQIAGGVVLAWSFRMVVDGAMFNQFIWYREVYVNMKEVFMSSVF 180

Db 121 ALHCHCSSPGSLQIAGGVVLAWSFRMVVDGAMFNQFIWYREVYVNMKEVFMSSVF 180

QY 181 MRGRHLVILRLWDGVHGVSVVPAMSGYSALHCGILNNIVLVCCSYCADLSEIRVRCAR 240

Db 181 MRGRHLVILRLWDGVHGVSVVPAMSGYSALHCGILNNIVLVCCSYCADLSEIRVRCAR 240

QY 241 RTRRLMLRAVRITAEETAMLYSCRTERRRQQOIFRALLQHRPILMHYDSTPM 294

Db 241 RTRRLMLRAVRITAEETAMLYSCRTERRRQQOIFRALLQHRPILMHYDSTPM 294

RESULT 2

S10867

early E4 34K protein - human adenovirus 12

C:Species: Mastadenovirus h12 (human adenovirus 12)

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999

C:Accession: S10867; S33952

R:Hogenkamp, T.; Esche, H.

Query Match 20.8%; Score 331; DB 1; Length 153;
Best_Local_Similarity 38.5%; Pred.No. 7.2e-24;
Matches 80; Conservative 10; Mismatches 42; Indels 76; Gaps 4;
ID: 1-MTSGVDFGWLRLRTRSLRRTPYGRDLRPFETETRILEDHPLPECNLTLMHNVS 60

```

5 RESULT
  QOBE4
  C:Species: human herpesvirus 4, Epstein-Barr virus
  C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 23-Aug-1997
  C:Accession: C93065; A03743
  C:Bankier, A.T.; Deininger, P.J.; Farrell, P.J.; Barrell, B.G.
  Mol. Biol. Med. 1, 21-45, 1983
  A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr vi
  A:Reference number: A93065; MUID:85035713
  A:Accession: C93065
  A:Molecule type: DNA
  A:Residues: 1-191 <BAN>

```

R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H
Nature 310, 207-211, 1984
A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A:Reference number: A03794; MUID:84270667
A:Contents: annotation; protein coding region
C:Superfamily: human herpesvirus 4 BHRF1 protein
C:Keywords: transmembrane protein

Query Match 5.6%; Score 88.5; DB 1; Length 191;
Best Local Similarity 23.3%; Pred. No. 0.45;
Matches 49; Conservative 18; Mismatches 66; Indels 77; Gaps 14;
QY 11 TLRPTRSLRSTPYSDRLPFFETATILEDHPLPECC-----NTLMMHN 58
DB 24 TLHPVLELAARETP-----RLSPEDT-----VLRVHVLEELIERNSETFTTNRFTIHT 76
QY 59 VS-----YVRLGPGSVGFTLIQWVPMVDMVLTREELVILKCMHVC--LCCA 104
DB 77 EHVLDLDFNSVFLIEFRKGP--SIGRAL--AWNA-W-----CMHACRTLCN 118
QY 105 N-----IDMTSMIT-----HGVESWALHCHCSPGSLQIAGGVVLSFWMVVDGAMFN 155
DB 119 QSTPYVVDLSVRGNLEAGLDGN--IHQQ-----GG-----WSTLIEDNIPGS 161
QY 156 QRIWTRVNVNMPREVMFMSVFMGRH 185
DB 162 RRFESWTLFLAGTLLSLVICSYLFSRGRH 191

RESULT 6
T32644
hypothetical protein F56A11.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
A:Accession: T32644
R:Gatting, S.; Goela, D.; Harper, M.
submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid F56A11.
A:Reference number: 221204
A:Accession: T32644
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-494 <GAT>
A:Cross-references: EMBL:AF038619; PIDN:AB92076.1; GSPDB:GNO0022; CESP:F56A11.4
A:Experimental source: strain Bristol N2; clone F56A11
C:Genetics:
A:Gene: CESP:F56A11.4
A:Map position: 4
A:Introns: 123/3; 235/2; 297/3; 337/3; 401/2; 471/2

Query Match 5.5%; Score 87; DB 2; Length 494;
Best Local Similarity 20.2%; Pred. No. 1.7;
Matches 51; Conservative 40; Mismatches 96; Indels 66; Gaps 10;
QY 36 ETATILEDHPLPECNITMHNVSVRGLPCSVGFTLIQWVPMVDMVLTREELVILRK 95
DB 53 EINDTCAEDP---PQADAFRMILIVIGVVCSLGI-----VLNTFLLSLRLDVFERS 103
QY 96 CMHVCLCCANIDMTS-----NMHIG-----YESWALHCHCSPGSLQIAGGQ 139
DB 104 NILYFLACLDLILVLCFMLIPASLVWDYFRVELLYTCWHFYIKYVS-----TVGQ 156
QY 140 VLASFWRVVDGAMFNQRTWYREVNVNMPKVFMS-----SVPMRGRHLIYLRWTD 194
DB 157 VLIAASTLLIVAASFERYICLSKSSIOFSPQRFLISIVGACALPKWSVFEEL--- 214
QY 195 GHVGSVPVAMSGYSAHCGLLNNIVLCCSYCADLSEIRVRCARTRMLMRAVRIIA 254
DB 214 -----QSLPCHPFPQNLRL-----DLSEI-----TRRTKLYIGKLIFS 248
QY 255 EE-TTAMLYSCRT 266

Db 249 ESYKTIWMFWCRS 261
RESULT 7
A70390
formate dehydrogenase (EC 1.2.1.2) gamma chain - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A70390
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.B.
V. Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: A70390
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-211 <AAQ>
A:Cross-references: GB:AE000720; NID:92983529; PIDN:AA07109.1; PID:92983534; GB:AE00
A:Experimental source: strain VF5
C:Genetics:
A:Gene: fdol
C:Superfamily: formate dehydrogenase gamma chain
C:Keywords: heterotrimer; NAD; oxidoreductase; transmembrane protein

Query Match 5.4%; Score 85.5; DB 1; Length 211;
Best Local Similarity 23.4%; Pred. No. 0.95;
Matches 36; Conservative 20; Mismatches 35; Indels 63; Gaps 7;
QY 64 GLPCSVGFTLIQWVPMVDMVLTREELVILRK-----M 97
DB 67 GVFSIGAVLM---ILKWARDVITREDIVMLTKIKAYISGRHEELPEVKYNAGKLYA 123
QY 98 HVCCLCANIDMTSMIT-----HGVESWALHCH-----CSSPCS 131
DB 124 WVFSLAIVFFLTGILMWFENFMSGLVRVSVVTHEITFIAGFTIIHYMATVGVPGS 183
QY 132 LQCIAGGQVLASFWRVVDGAMFNQRFIWRV 165
DB 184 IWGMIGGKVSATW-----AKFHHP-KWYREV 209

RESULT 8
T24098
hypothetical protein R09H10.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24098
R:Basham, V.
submitted to the EMBL Data Library, July 1996
A:Reference number: Z19840
A:Accession: T24098
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1603 <WII>
A:Cross-references: EMBL:Z77134; PIDN:CAB00874.1; GSPDB:GNO0022; CESP:R09H10.5
A:Experimental source: clone R09H10
C:Genetics:
A:Gene: CESP:R09H10.5
A:Map position: 4
A:Introns: 18/1; 57/3; 75/2; 120/1; 168/2; 194/1; 302/2; 348/3; 392/3; 442/1;

Query Match 5.1%; Score 82; DB 2; Length 1603;
Best Local Similarity 21.9%; Pred. No. 17;
Matches 44; Conservative 29; Mismatches 82; Indels 46; Gaps 11;
QY 9 GMTLRPTRSLRSTPYSDRLPFFETATILEDHPLPECNITMHNVSV-----R 63
DB 972 GFSPDPSPDFSRDLTYATHQOPVHL-----SSTLQSPDSIDVTSYNNNNIATSLFTR 1028

QY 64 GLPCSVGFTLLQEW-----VVPWMLVRLREELVILKCMHYC-----LCANIDIM---T 110
Db 1029 TSGKYEYIMNOEYTCQASGDTFTMETITINVTQRAVCSLDICINGGIMLGN 1089
QY 111 SMINGYESWALHCH---CSPGSL-----OC---IAGQVSLASFRWVVDGMFNQRFIW 160
Db 1089 CQCVNGYTS--LHCEVPTCNGSVDFKQCPSTYDG-----DSQYITCTNTW 1135
QY 161 YREVYNYMPKEVFMSSVPM 181
Db 1136 --NFVETHDPREYNFOQIVFV 1154

RESULT 9
IS1200
rhodopsin - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: IS1200; S31474; S31473
R:Saha, M.S.; Grainger, R.M.
A:Title: Early opsin expression in xenopus embryos precedes photoreceptor differentiation
A:Reference number: IS1200; MUID:93287804
A:Accession: IS1200
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-354 <SAH>
A:Cross-references: GB:S62229; NID:9386187; PID:AB27128.1; PID:9386188; EMBL:L04692; N
R:Knox, B.E.; Scalzetti, L.C.; Batni, S.; Wang, J.Q.
submitted to the EMBL Data Library, December 1992
A:Description: Molecular Cloning of the abundant rod opsin and transducin from xenopus l
A:Reference number: S31474
A:Accession: S31474
A:Molecule type: mRNA
A:Residues: 1-106, 'Q', 108-136, 'M', 138-240, 'A', 242-354 <KNO>
A:Cross-references: EMBL:L07770; NID:9214734; PID:AAC42232.1; PID:9214735
C:Keywords: chromoprotein; eye; G protein-coupled receptor; glycoprotein; lipoprotein; R
E:74-96/Domain: transmembrane #status predicted <TM1>
E:114-133/Domain: transmembrane #status predicted <TM2>
E:153-175/Domain: transmembrane #status predicted <TM3>
E:203-230/Domain: transmembrane #status predicted <TM4>
E:253-276/Domain: transmembrane #status predicted <TM5>
E:285-309/Domain: transmembrane #status predicted <TM6>
E:315/Binding site: carbohydrate (Asn) (covalent) #status predicted <TM7>
E:296/Binding site: retinal (Lys) (covalent) #status predicted

Query Match 5.1%; Score 81.5; DB 1; Length 354;
Best Local Similarity 20.0%; Pred. No. 3.9; Mismatches 94; Indels 117; Gaps 13;
Matches 61; Conservative 33;

QY 5 GVPFG-MTLRPTSRSLRRTYPSRDLPPFTETRAILEDHPLPCECNLTLMHNV--- 60
Db 51 GLPINFMLFTVIOHKLRTPLNLL-----NLVFANHEWVLGGFTVMTSHGY 102
QY 60 -----SYVRGLPCSVGFTLLQEWVVPWMLVRLREELVILKCMHYCCLCANIDIMTSM 112
Db 103 FIFGPTGCIIEGF-----FATLGEVALMSLVLAVERIV-----VCKPMANFRFGNH 152
QY 113 MHGVE-SWALHCHCSPGSLQCIAGQVSLASFRWVVDG-----AMFNOR 157
Db 153 AIMGVATWIALSCAAP-----PLFGWSRVIPGMCQSCGVDDYTLKPVNNS 202
QY 158 FIWYREVYNYMPKEVFMSSVPMRGRHL----- 187
Db 203 FVIYMFVHTIPLIVF-----FCYGRLLCTKEAAQOQESLTTQAKKEVTRVWIMV 258
QY 187 -----IYLRWDVGHVGS-----VVPAMSGYSALHCG-----ILNN-----I 219
Db 259 VFELICWPIYAVFYIFTHQSGNFGVPMTPVPAFAKSAIYNPIVILNKQFNCLI 318

QY 220 VLQCC 224
Db 319 TTLCC 323

RESULT 10

T16796

hypothetical protein T04A6.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 24-Nov-1999

C:Accession: T16796

R:Vaudin, M.; Anderson, K.

submitted to the EMBL Data Library, April 1994

A:Description: The sequence of C. elegans cosmid T04A6.

A:Reference number: Z18579

A:Accession: T16796

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-456 <VAU>

A:Cross-references: EMBL:U00042; NID:9470332; PID:9470335; PIDN:AAA50674.1; CESP:T04A

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:T04A6.3

A:Introns: 23/3; 76/3; 104/3; 232/1; 250/2; 311/2; 340/3; 374/1; 402/3

C:Superfamily: Caenorhabditis elegans hypothetical protein T04A6.3

Query Match

Best Local Similarity 19.5%; Score 79.5; DB 2; Length 456;

Matches 54; Conservative 30; Mismatches 72; Indels 121; Gaps 13;

QY 8 FGMLRPTSRSL-----SRTPYSRDLPPFTETRAILEDHPLPCECNLTLMHNV 60
Db 50 FSVTLNHTPRTALFPWNKSFQTPITAEAP-----ISMLP-CGT----- 89

QY 61 YVRGLPCSVGFTLLQEWVVPWMLVRLREELVILKCMHYCCLCANIDIMTSMTHGESW 120
Db 89 -----CRYN-----LPWVVIIEGKV-----AANF-----SW 111

QY 121 ALHCHCSPGSLQCIAGQVSLASFRWVVDGAMFNQRFIYREVNV----- 167
Db 112 FNNDCVKSQGLVRLGPGQVGVGHEYPPEPANFDAGKIWKPVVNNHKGDISFGV 171

QY 167 YNPKVFMSSVPMRGRHLIYLRWDVGHVSVVPAMSGY-----SALHCGILNNI 219
Db 172 VNLPCQKQIFAKVDVRNE-----SYGNQKGKHSARGPACASSY 211

QY 220 VYLCCSYCADL-----SEIRVRCARR 241
Db 212 SILCKLQDNTSRDRKNTYSVLSDDDNDRLENCGR 248

RESULT 11

T26242

hypothetical protein W06F12.2a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T26242

R:Lloyd, C.

submitted to the EMBL Data Library, December 1996

A:Reference number: Z20179

A:Accession: T26242

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-562 <WIL>

A:Cross-references: EMBL:Z83244; PIDN:CAB54314.1; GSPDB:GN00021; CESP:W06F12.2a

A:Experimental source: clone W06F12

C:Genetics:

A:Gene: CESP:W06F12.2a

A:Map position: 3

A:Introns: 23/3; 75/1; 153/3; 201/1; 283/1; 305/3; 332/2; 555/3

us-09-214-478-2.rpt

Mon Jun 26 09:10:37 2000

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 07-Dec-1999

C:Accession: B58851
 R:Aranson, U.; Gullberg, A.; Widegren, B.
 J. Mol. Evol. 33, 556-568, 1991
 A:Title: The complete nucleotide sequence of the mitochondrial DNA of the fin whale.
 A:Reference number: A58850; MUID:92139449
 A:Accession: B58851
 A:Molecule type: DNA
 A:Residues: 1-459 CARC
 A:Cross-references: GB:X61145; NID:g12772; PIDN:CAA43448.1; PID:g12780
 C:Genetics:
 A:Genome: mitochondrion
 A:Genetic code: SGC1
 A:Superfamily: NADH dehydrogenase (ubiquinone) chain 4
 C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation

Query Match 5.0%; Score 79; DB 2; Length 459;
 Best Local Similarity 19.9%; Pred. No. 8.8; Indels 74; Gaps 9;
 Matches 37; Conservative 29; Mismatches 50; Indels 74; Gaps 9;

QY 81 WDMVLR-----EELVLRKCMHVCCLCCANIDIMTS-----MMIHGYESMAL 122
 DB 267 WGMIMTSSICLRQDLKSLAYSSVSHMALVIAAILIQTWPSTMGATAMIAHGLTSSML 326
 QY 123 HCHCSS-----PSLQCIAGGVLASFWRVVDGAMFNQRFIWEVYNNMP 170
 DB 327 FCLANSYERHSRTMILPGLQVFL--PLMASWLLA-----SLTNALP 370
 QY 171 K-----EVMFMSSVF-----MRGRHLIYLRM-----YGHVGSVVPAM 204
 DB 371 PTNLIGELVMSVFSNSNTILLMGNTIVITALYSLYMLTQKGHTHHNNITPSF 430
 QY 205 SFGYS--ALH 212
 DB 431 TREHALMALH 440

RESULT 14

S36506
 E2 protein - human papillomavirus type 30
 C:Species: human papillomavirus type 30
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999
 C:Accession: S36506
 R:Delius, H.; Hofmann, B.
 Submitted to the EMBL Data Library, August 1993
 A:Description: Primer-directed sequencing of human papillomavirus types.
 A:Reference number: S36469
 A:Accession: S36506
 A:Molecule type: DNA
 A:Residues: 1-378
 A:Cross-references: EMBL:X74474; NID:g396973; PIDN:CAA52546.1; PID:g396977
 C:Superfamily: papillomavirus E2 protein
 C:Keywords: DNA binding; early protein; transcription regulation

Query Match 4.9%; Score 78; DB 2; Length 378;
 Best Local Similarity 21.6%; Pred. No. 8.9; Indels 68; Gaps 14;
 Matches 50; Conservative 27; Mismatches 86; Indels 68; Gaps 14;

QY 74 IOEYVVPDMLTREELVLRK-----CMHVCLCCANIDIMTSMMI----- 115
 DB 26 IEDHYVYKAV--RHENVVLYKARQNNITKLHQQVPCVQCKAKACVAIEIQMALESLEY 83
 QY 115 ---HGYESWALHCHCSS---PGSLQCI-AGGVLASFWRVVDGAMFN--ORFIWEVY 165
 DB 84 KTEYKVEEWTLDKVCENMMHTAPKQCFKSKGRIEVWF-----DGKDNREYVYVW--QWV 137
 QY 166 NY-----NMPKYMFMSSVEMRGRHLIYLRWYDGHVGSVVPAMSGYSA---LHCGI 215
 DB 138 YCGDNGWTKVPSVDYKGIYYVHDGNKYVYTDND-----EAVKYGKGTWEVHMG- 190
 QY 216 LNNIVLCCSYCADLSIEVRCCARRRMLMLRAVRIIAEETMAYLSCT 266

Query Match 5.0%; Score 79.5; DB 2; Length 562;
 Best Local Similarity 23.4%; Pred. No. 9.8; Indels 55; Gaps 9;
 Matches 39; Conservative 25; Mismatches 48; Indels 55; Gaps 9;

QY 100 CLCCANID-----IMTSMIHGYESWALHCHC---SPGSLQCC----- 135
 DB 6 CCCSSKDAITIGISLVYALASLLFLQWGTGLN-HCRVVTMAQSNLQCEWDCPCVGA 64
 QY 135 -----IAGG-----OVL-----ASWRMVVDGAMFNQRFIWEVYNNMPKVEVFMSSVF 180
 DB 65 SSKRTSALIEGLFVQVLCILIAFFLVFASAAIYGIHTWSKYLW---WPWFVPMLSLIL 121
 QY 181 MRGRHLIYLRWYDGHVGSVVPAMSGYSAHCGILNNIVLCCSYC 227
 DB 122 AT---LAYCIMWCGDVRVYLAIT-----IIEIIVFINIYC 156

RESULT 12

S40688
 opsin, rod - Pomatoschistus minutus
 C:Species: Pomatoschistus minutus
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: S40688
 R:Archer, S.N.; Lythgoe, J.N.; Hall, L.
 Proc. R. Soc. Lond. B Biol. Sci. 248, 19-25, 1992
 A:Title: Rod opsin cDNA sequence from the sand goby (Pomatoschistus minutus) compared with other teleost sequences.
 A:Reference number: S40688; MUID:92396703
 A:Accession: S40688
 A:Molecule type: mRNA
 A:Residues: 1-352 CARC
 A:Cross-references: EMBL:X62405; NID:g433817; PIDN:CAA44275.1; PID:g433818
 C:Superfamily: vertebrate rhodopsin
 C:Keywords: chromoprotein; eye; G protein-coupled receptor; glycoprotein; lipoprotein; F; 37-61/Domain: transmembrane #status predicted <TM2>
 F; 74-96/Domain: transmembrane #status predicted <TM2>
 F; 114-140/Domain: transmembrane #status predicted <TM3>
 F; 153-173/Domain: transmembrane #status predicted <TM4>
 F; 203-230/Domain: transmembrane #status predicted <TM5>
 F; 253-276/Domain: transmembrane #status predicted <TM6>
 F; 286-309/Domain: transmembrane #status predicted <TM7>
 F; 110-187/Disulfide bonds: #status predicted
 F; 296/Binding site: retinal (Lys) (covalent) #status predicted

Query Match 5.0%; Score 79; DB 1; Length 352;
 Best Local Similarity 19.9%; Pred. No. 6.6; Indels 76; Gaps 13;
 Matches 51; Conservative 31; Mismatches 68; Indels 76; Gaps 13;

QY 55 TMHNVSYRGLPCSV-GP--TLIOEYVVPDMLTREELVLRKCMHVCCLCCANIDIMTS 111
 DB 98 SMHGTVFLGRGCVGEGFATLGE-IALMSLVLAVERVWV-----VCKPISNFRFTEN 151
 QY 112 MHIHGYE-SWALHCHCSSPGSLQCIAGGVLASFWRVVDG-----AMFNQ 156
 DB 152 HAIMGVAFSWIAATCAVP-----PLVGSRYIPEGMQCSGVDYITRAEGENNE 201
 QY 157 RFTWYREVYNNMPKYMFMSSVEMRGRHLIYLR-----FCYGRLLCAVKEAQAQSEPTORAEVTRMVIIM 257
 DB 202 SFVIYMFIVHFLAPLVIF-----FCYGRLLCAVKEAQAQSEPTORAEVTRMVIIM 257
 QY 191 -----LWV-----DGHVGSVVPAMSGYSAHCGILNNIVLCCS--- 226
 DB 258 VIGFLSWLPYASVAVYITHQGTGFLPMTIPAFKSSALY-----NPMIYICMKNQF 313
 QY 226 -YCADLSEIVRCAR 240
 DB 314 RHCM-----ITTLCCGK 325

RESULT 13

B58851
 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - finback whale mitochondrion
 C:Species: mitochondrion Balaenoptera physalus (finback whale, common rorqual)

```

Db 190 NESI-----YCPD-----SVSSTLRNVSPVETWVEYNTYNYQPT 226

RESULT 15
T11502
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - dog mitochondrion (fragment)
C:Species: mitochondrion Canis lupus familiaris (dog)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 07-Dec-1999
C:Accession: T11502
R:Kim, K.S.; Lee, S.B.; Jeong, H.W.; Ha, J.H.
Mol.: Phylogenet. Evol. 10, 210-220, 1998
A:Title: The complete nucleotide sequence of the domestic dog (Canis familiaris) mitochondrion
A:Reference number: Z17276; MUID:99097286
A:Accession: T11502
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-459 <KIM>
A:Cross-references: EMBL:U96639; NID:g4154170; PID:g4154180; PIDN:AAD04772.1
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC1
A:Note: ND4
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

```

```

Query Match 4.9%; Score 77.5; DB 2; Length 459;
Best Local Similarity 24.1%; Pred. No. 12;
Matches 26; Conservative 14; Mismatches 37; Indels 31; Gaps 5;

QY 40 TILEDHPLLPCEWTLMHNSVYVGRGLPCSVGFTLIQEWVYP-----WDMVLTREELVLRK 95
|::|||
|::|||
DB 153 TLMGSLPLVA-----LLYHFMGSLNFMQVWQIPLPNSWSNIF----- 195

QY 96 CMHVCCLCANTIDMTSMHIGYESWALHCHCSPGSLQCIAGQVLA 143
|::|||
|::|||
DB 195 ---LWLACM-NAFVNMPLYGLHLWLPKAVHAP-----IAGSMVLA 233
|::|||
|::|||

```

Search completed: June 23, 2000, 10:05:43.
Job time: 1748 sec

Mon Jun 26 09:10:39 2000

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2000, 10:03:39 ; Search time 13.99 Seconds
(without alignments)
640.012 Million cell updates/sec

Title: US-09-214-478-2
Perfect score: 1594
Sequence: 1 MTSGVPGMTLRTRSLRS.....RALQHRPILMHYDSTPM 294

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 83857 seqs, 30454973 residues

Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_38.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1594	100.0	294	1 E434_ADE02	P03239 human adeno
2	990	62.1	292	1 E434_ADE09	P89083 human adeno
3	874.5	54.9	291	1 E434_ADE12	P36710 human adeno
4	706	44.3	257	1 E434_ADE40	P64865 human adeno
5	331	20.8	153	1 E417_ADE02	P03238 human adeno
6	253.5	15.9	259	1 E434_ADECT	P87568 canine adeno
7	252.5	15.8	265	1 E434_ADECR	P96690 canine adeno
8	250.5	15.7	265	1 E434_ADECC	P65962 canine adeno
9	186.5	11.7	289	1 E434_ADEM1	P23125 mouse adeno
10	91.5	5.7	643	1 UROM_BOVIN	P48733 bos taurus
11	88.5	5.6	191	1 EAR_EBV	P03182 epstein-bar
12	81.5	5.1	334	1 OPSD_XENLA	P29403 xenopus lae
13	79	5.0	249	1 OPSD_LIMBE	O42427 limnocottus
14	79	5.0	352	1 OPSD_POMMI	P35403 ponatoschis
15	79	5.0	354	1 OPSD_ZEUFA	O42604 zeus faber
16	78	5.0	459	1 NU4M_BALPH	P24975 balaenopter
17	78	4.9	378	1 VE2_HPV30	P36790 human papil
18	77.5	4.9	459	1 NU4M_CANFA	P35350 brachydanio
19	77	4.8	289	1 OPSD_BATMU	P03910 bos taurus
20	76.5	4.8	354	1 OPSU_BRARE	P48016 felis silve
21	76.5	4.8	459	1 NU4M_BOVIN	P48016 felis silve
22	76.5	4.8	459	1 NU4M_FELCA	P48016 felis silve
23	76.5	4.8	598	1 NTG1_MUSCO	P48016 felis silve
24	76	4.8	504	1 DGTG_ECOLI	P15723 escherichia
25	76	4.8	504	1 DGTG_ESCFE	P15723 escherichia
26	76	4.8	518	1 PELL_YEAST	O59435 escherichia
27	75.5	4.7	458	1 ALG3_YEAST	P25178 saccharomyc
28	75	4.7	349	1 OPSD_MYRBE	P38179 saccharomyc
29	74	4.6	196	1 HTGA_ECOLI	P79798 myristicis
30	74	4.6	350	1 IL8A_GORGO	P28697 escherichia
31	74	4.6	1255	1 ERB2_HUMAN	P55919 gorilla gor
32	73.5	4.6	446	1 NU4M_DROYA	P04626 homo sapien
33	73.5	4.6	475	1 YKGF_ECOLI	P07707 drosophila
34	73.5	4.6	604	1 KRAA_RAT	P77536 escherichia
					P14056 rattus norv

35	73.5	4.6	606	1 KRAA_HUMAN	P10398 homo sapien
36	73.5	4.6	606	1 KRAA_PIG	O19004 sus scrofa
37	73	4.6	289	1 OPSD_PARKN	O42452 paracottus
38	73	4.6	348	1 OPSD_TRIMA	O62796 trichechus
39	73	4.6	391	1 GPD1_YEAST	Q00053 saccharomyc
40	73	4.6	504	1 DGTG_SHIBO	Q09827 shigella bo
41	73	4.6	515	1 GLCM_MOUSE	P17439 mus musculu
42	72.5	4.5	352	1 C5AR_CANFA	P30992 canis famli
43	72.5	4.5	459	1 NU4M_BALMU	P41298 balaenopter
44	72.5	4.5	459	1 NU4M_SHEEP	O78755 ovis aries
45	72	4.5	289	1 OPSD_ABYKO	O42294 abyssocottu

ALIGNMENTS

RESULT 1

ID	E434_ADE02	STANDARD;	PRT;	294 AA.
AC	P03239;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	EARLY E4 34 KD PROTEIN.			
OS	Human adenovirus type 2.			
OC	Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 82059444.			
RA	Herisse J., Rigolet M., Dupont de Dinechin S., Galibert F.;			
RT	"Nucleotide sequence of adenovirus 2 DNA fragment encoding for the			
RL	carboxylic region of the fiber protein and the entire E4 region.";			
CC	Nucleic Acids Res. 9:4023-4042(1981);			
CC	-!- MISCELLANEOUS: FOUND IN ASSOCIATION WITH E1B 55 KD PROTEIN			
CC	-!- SIMILARITY: BELONGS TO THE ADENOVIRUS E4 30 TO 34 KD PROTEIN			
CC	FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	at the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; J01917; -; NOT ANNOTATED_CDS.			
DR	PIR; A03805; Q4ADC2.			
KW	Early protein.			
SQ	SEQUENCE 294 AA; 34116 MW; F60C83A38240BE0C CRC64;			

Query Match 100.0%; Score 1594; DB 1; Length 294;
Best Local Similarity 100.0%; Pred No. 6.3e-148;
Matches 294; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MTTSGVPGMTLRTRSLRSRTTYSRDLPPPTETETATILEDHPLPNCNTLTHNVS	60
DB	1	MTTSGVPGMTLRTRSLRSRTTYSRDLPPPTETETATILEDHPLPNCNTLTHNVS	60
QY	61	YVRLPCSVGTLTQEWVVPDWMVLTREELVILKRMHVCLCCANIDIMTSMIHYESW	120
DB	61	YVRLPCSVGTLTQEWVVPDWMVLTREELVILKRMHVCLCCANIDIMTSMIHYESW	120
QY	121	ALRHCSPPGSLQCIAGGQVLASWFMVVDGAMFNQRFITWREVNTNMPKEVFMMSVF	180
DB	121	ALRHCSPPGSLQCIAGGQVLASWFMVVDGAMFNQRFITWREVNTNMPKEVFMMSVF	180
QY	181	MRGRHLIYLRWDHGVSVVPAMSGYSAHLCGLTNIVVLCSCYCADLSEIRVRCAR	240
DB	181	MRGRHLIYLRWDHGVSVVPAMSGYSAHLCGLTNIVVLCSCYCADLSEIRVRCAR	240
QY	241	RTFRMLRAVRITAEETTMALYSCRTRRRROOFIRALQHRPILMHYDSTPM	294

```

Db 241 RTRMLRAVRIIAETATMLYSCRTERROQFIRALLQHRPILMHVDYDSTPM 294
RESULT 2
E434_ADE09
ID E434_ADE09 STANDARD; PRT; 292 AA.
AC P89083;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE EARLY E4 34 KD PROTEIN.
OS Human adenovirus type 9.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96327849.
RA Javner R., Shenk T.;
RT "Mammary tumors induced by human adenovirus type 9: a role for the
RL viral early region 4 gene."
CC Breast Cancer Res. Treat. 39:57-67(1996).
CC -!- SIMILARITY: BELONGS TO THE ADENOVIRUS E4 30 TO 34 KD PROTEIN
FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; S82508; AAB37508.1;
CC Early protein.
CC
CC Query Match 62.1%; Score 990; DB 1; Length 292;
CC Best Local Similarity 63.7%; Pred. No. 3.8e-89;
CC Matches 174; Conservative 45; Mismatches 52; Indels 2; Gaps 1;
CC
CC QY 21 RRTPYSDRLPPPTETRAILEDHPLPECNLTAMHNSVYVGLPCSVGFTLIQEWVP 80
CC Db 11 RHPPYRARLPDSDETRASLEQHPFLPCDHDYHNSVSSVGLPCAAGFTLLQEPFVP 70
CC QY 81 WDMVLTREELVILKRCMHVCLCCANIDIMTSMIHGYESWALHCHCSPGSLQCIAGSQV 140
CC Db 71 WDMILTPEEIKILKRCMSYCLCPATLDLVRAQMVSGYERWILHCHCSPGSLQCRAGGTL 130
CC QY 141 LASWFRMVVDGAMFNORFTWREVNNPKPEVMFMSVFMGRHLLIYLRWYDGHVGSV 200
CC Db 131 LAVWFRVYVGMFNORFTWREVNNPKPEIMYGMVFMGRHLLIYCRWYDGHVGSV 190
CC QY 201 VPAMSGYSAHCGILNNIVLVCCSYCADLSEIRVRCARRTRMLRAVRIIAET--T 258
CC Db 191 IPNMSFGWSALNYGLNNVIMWCCTYCENMAEIRMRCCARRTRMLKAVGIIVRETCDP 250
CC QY 259 AMLYSCRTERROQFIRALLQHRPILMHVDYS 291
CC Db 251 DPICSTRTEPRQRLRALMRHRPILFSEYES 283
RESULT 3
E434_ADE12
ID E434_ADE12 STANDARD; PRT; 291 AA.
AC P36710;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE EARLY E4 34 KD PROTEIN.
OS Human adenovirus type 12.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE; 94076430.
RA Sprengel J., Schmitz B., Heuss-Neitzel D., Zock C., Doerfler W.;
RT "Nucleotide sequence of human adenovirus type 12 DNA: comparative
RT functional analysis."
RL J. Virol. 68:379-389(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90272430.
RX Hogenkamp T., Esche H.;
RT "Nucleotide sequence of the right 10% of adenovirus type 12 DNA
RT encoding the entire region E4."
RL Nucleic Acids Res. 18:3065-3066(1990).
CC -!- SIMILARITY: BELONGS TO THE ADENOVIRUS E4 30 TO 34 KD PROTEIN
FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X73487; CAAS1901.1;
CC EMBL; X51800; CAB57854.1;
CC PIR; S33952; S33952.
CC PIR; S10867; S10867.
CC Early protein.
CC CONFLICT 228 229 CA -> WP (IN REF. 2).
CC SEQUENCE 291 AA; 34383 MW; CDIA07D37FBB5E96 CRC64;
CC
CC Query Match 54.9%; Score 874.5; DB 1; Length 291;
CC Best Local Similarity 55.7%; Pred. No. 6.5e-78;
CC Matches 151; Conservative 52; Mismatches 67; Indels 1; Gaps 1;
CC
CC QY 21 RRTPYSDRLPPPTETRAILEDHPLPECNLTAMHNSVYVGLPCSVGFTLIQEWVP 79
CC Db 10 RLAPYKYLQPPCEQSKATYSTNSLWPCNSLTAMHNSVSEVRGIPSCVFTVLEWPI 69
CC QY 80 PDMVLTREELVILKRCMHVCLCCANIDIMTSMIHGYESWALHCHCSPGSLQCIAGSQ 139
CC Db 70 PDMILTPEEIKILKRCMSYCLCPATLDLVRAQMVSGYERWILHCHCSPGSLQCSAGM 129
CC QY 140 VIASFVMDVGMFNORFTWREVNNPKPEVMFMSVFMGRHLLIYLRWYDGHVGSV 199
CC Db 130 LLGRFKNVAVGALINKRCFWREVNNPKPEVMFMSVFMGRHLLIYLRWYDGHVGSV 189
CC QY 200 VVPAMSGYSAHCGILNNIVLVCCSYCADLSEIRVRCARRTRMLRAVRIIAET 259
CC Db 190 ALEKVSFGWSAFNYGILNNVIMWCCTYCENMAEIRMRCCARRTRMLKAVGIIVRETCDP 249
CC QY 250 MLYSCRTERROQFIRALLQHRPILMHVDYD 290
CC Db 250 PLKISRHERYRQQLKGLIMHRAILFGDYN 280
RESULT 4
E434_ADE40
ID E434_ADE40 STANDARD; PRT; 257 AA.
AC Q64865;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE EARLY E4 30 KD PROTEIN.
OS Human adenovirus type 40.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=DUGAN.
RX MEDLINE; 94087748.
RA Davison A.J., Telford E.A., Watson M.S., McBride K., Mautner V.;
RT "The DNA sequence of adenovirus type 40."

```

Query Match	Best Local Similarity	Score	DB 1:	Length	DB 2:	Score	DB 3:	Length
Query Match	Best Local Similarity	Score	DB 1:	Length	DB 2:	Score	DB 3:	Length
Matches	80; Conservative	10; Mismatches	42; Indels	76; Gaps	4;			
1	1	153	EARLY 17 KD PROTEIN.					
2	2	153	EARLY 10 KD PROTEIN.					
3	3	153	EARLY 10 KD PROTEIN.					
4	4	153	EARLY 10 KD PROTEIN.					
5	5	153	EARLY 10 KD PROTEIN.					
6	6	153	EARLY 10 KD PROTEIN.					
7	7	153	EARLY 10 KD PROTEIN.					
8	8	153	EARLY 10 KD PROTEIN.					
9	9	153	EARLY 10 KD PROTEIN.					
10	10	153	EARLY 10 KD PROTEIN.					
11	11	153	EARLY 10 KD PROTEIN.					
12	12	153	EARLY 10 KD PROTEIN.					
13	13	153	EARLY 10 KD PROTEIN.					
14	14	153	EARLY 10 KD PROTEIN.					
15	15	153	EARLY 10 KD PROTEIN.					
16	16	153	EARLY 10 KD PROTEIN.					
17	17	153	EARLY 10 KD PROTEIN.					
18	18	153	EARLY 10 KD PROTEIN.					
19	19	153	EARLY 10 KD PROTEIN.					
20	20	153	EARLY 10 KD PROTEIN.					
21	21	153	EARLY 10 KD PROTEIN.					
22	22	153	EARLY 10 KD PROTEIN.					
23	23	153	EARLY 10 KD PROTEIN.					
24	24	153	EARLY 10 KD PROTEIN.					
25	25	153	EARLY 10 KD PROTEIN.					
26	26	153	EARLY 10 KD PROTEIN.					
27	27	153	EARLY 10 KD PROTEIN.					
28	28	153	EARLY 10 KD PROTEIN.					
29	29	153	EARLY 10 KD PROTEIN.					
30	30	153	EARLY 10 KD PROTEIN.					
31	31	153	EARLY 10 KD PROTEIN.					
32	32	153	EARLY 10 KD PROTEIN.					
33	33	153	EARLY 10 KD PROTEIN.					
34	34	153	EARLY 10 KD PROTEIN.					
35	35	153	EARLY 10 KD PROTEIN.					
36	36	153	EARLY 10 KD PROTEIN.					
37	37	153	EARLY 10 KD PROTEIN.					
38	38	153	EARLY 10 KD PROTEIN.					
39	39	153	EARLY 10 KD PROTEIN.					
40	40	153	EARLY 10 KD PROTEIN.					
41	41	153	EARLY 10 KD PROTEIN.					
42	42	153	EARLY 10 KD PROTEIN.					
43	43	153	EARLY 10 KD PROTEIN.					
44	44	153	EARLY 10 KD PROTEIN.					
45	45	153	EARLY 10 KD PROTEIN.					
46	46	153	EARLY 10 KD PROTEIN.					
47	47	153	EARLY 10 KD PROTEIN.					
48	48	153	EARLY 10 KD PROTEIN.					
49	49	153	EARLY 10 KD PROTEIN.					
50	50	153	EARLY 10 KD PROTEIN.					
51	51	153	EARLY 10 KD PROTEIN.					
52	52	153	EARLY 10 KD PROTEIN.					
53	53	153	EARLY 10 KD PROTEIN.					
54	54	153	EARLY 10 KD PROTEIN.					

QY 199 SVPMASFGSALHGLINNIIVLCCSYCADLSEIRVRCARTRRLMRAVIAEETT 258
 DB 168 VLKSC-----TYNWLVLKCKSCSLQNYCALKNAWVRSYIVAKVEY-EKTP 215
 QY 259 AMY--SCRTERRRROQTRIRALLQHR 282
 DB 216 VILHQTTSKAEKRRQSLRQAMMGR 241
 RESULT 7
 E434_ADECR STANDARD: PRT: 265 AA.
 AC 096690;
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE EARLY E4 31 KD PROTEIN.
 OS Canine adenovirus type 1 (strain RI261).
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 CC [1]
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; Y07760; CAA69046.1; -
 CC DR Early protein.
 CC KW
 CC SEQUENCE 265 AA; 30737 MW; 14B3DC9DCB1D2E15 CRC64;
 Query Match 15.88; Score 252.5; DB 1; Length 265;
 Best Local Similarity 26.98; Pred. No. 2e-17;
 Matches 70; Conservative 43; Mismatches 96; Indels 51; Gaps 8;
 QY 51 CNTLMNVSVYRGLP--CSVGFLLIDWVPMVMYTRRELYI-----LRKM 97
 DB 11 CNAETSHVAVVAPAFPCNC-FALCLEIPLMDLLYRREKILFGFGTCNGAELILNS 69
 QY 98 HVCLOCANIDIMTSMIHGYSNALHCHGSSPSLOCIAGGOVLASFRRVYDGAENOR 157
 DB 70 HCLCLADA-----QMMOVHCHGSDSLQCLSTQVLAKEFLBEFVWGCVNKK 116
 QY 158 FIWREVVNTMPREVMSSVFRGRLIYLLMTWDGHVS-----VVPAMSEGY 208
 DB 117 YLMRREFVNSRDEINYGSIERNHYIFRSLFSTVHOACMLAIORCISPELGYVF 176
 QY 209 SALHGLINNIIVLCCSYCADLSEIRVRCARTRRLMRAV-----IIEETTAMLY 262
 DB 177 KSTY---NWVLVKKCKSCSLQNYCALKSCAFWVRSIIDRVLRVEKIPVLAHRTS--- 230
 QY 263 SCRTERRRROQTRIRALLQHR 282
 DB 230 --KAERKROTALQAMMYGR 247
 RESULT 8
 E434_ADECC STANDARD: PRT: 265 AA.
 AC 065962;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE EARLY E4 31 KD PROTEIN.
 OS Canine adenovirus type 1 (strain CL1).
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 CC [1]
 CC SEQUENCE FROM N.A.
 CC RA Campbell J.B., Zhao Y.;
 CC RU Submitted (Aug-1996) to the EMBL/Genbank/DBJ databases.
 CC RN [2]
 CC SEQUENCE OF 248-265 FROM N.A.
 CC RX MEDLINE; 91272490.
 CC RA Dragulev B.P., Sira S., Abouhaider M.G., Campbell J.B.;
 CC RT "Sequence analysis of putative E3 and fiber genomic regions of two
 CC strains of canine adenovirus type 1.";
 CC Virology 183:298-305(1991).
 CC -1- SIMILARITY: BELONGS TO THE ADENOVIRUS E4 30 TO 34 KD PROTEIN
 CC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; U55001; AB05452.1; -
 CC DR Early protein.
 CC KW
 CC SEQUENCE 265 AA; 30822 MW; 92606E5E30134103 CRC64;
 Query Match 15.78; Score 250.5; DB 1; Length 265;
 Best Local Similarity 26.98; Pred. No. 3.1e-17;
 Matches 70; Conservative 43; Mismatches 96; Indels 51; Gaps 8;
 QY 51 CNTLMNVSVYRGLP--CSVGFLLIDWVPMVMYTRRELYI-----LRKM 97
 DB 11 CNAETSHVAVVAPAFPCNC-FALCLEIPLMDLLYRREKILFGFGTCNGAELILNS 69
 QY 98 HVCLOCANIDIMTSMIHGYSNALHCHGSSPSLOCIAGGOVLASFRRVYDGAENOR 157
 DB 70 HCLCLADA-----QMMOVHCHGSDSLQCLSTQVLAKEFLBEFVWGCVNKK 116
 QY 158 FIWREVVNTMPREVMSSVFRGRLIYLLMTWDGHVS-----VVPAMSEGY 208
 DB 117 YLMRREFVNSRDEINYGSIERNHYIFRSLFSTVHOACMLAIORCISPELGYVF 176
 QY 209 SALHGLINNIIVLCCSYCADLSEIRVRCARTRRLMRAV-----IIEETTAMLY 262
 DB 177 KSTY---NWVLVKKCKSCSLQNYCALKSCAFWVRSIIDRVLRVEKIPVLAHRTS--- 230
 QY 263 SCRTERRRROQTRIRALLQHR 282
 DB 230 --KAERKROTALQAMMYGR 247
 RESULT 9
 E434_ADEM1 STANDARD: PRT: 289 AA.
 AC P23125;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PROBABLE EARLY E4 33 KD PROTEIN (ORF A/B PROTEIN).
 OS Mouse adenovirus type 1 (MAV-1).
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 CC [1]
 CC SEQUENCE FROM N.A.
 CC RX MEDLINE; 91082417.
 CC RA Ball A.O., Beard C.W., Villegas P., Spindler K.R.;
 CC RT "Early region 4 sequence and biological comparison of two isolates of
 CC mouse adenovirus type 1.";
 CC Virology 180:257-265(1991).

CC -1- MISCELLANEOUS: THIS PROTEIN IS THE PRODUCT OF TWO DIFFERENT MRNAS
 CC SPLICED FROM THE SAME PRIMARY TRANSCRIPT.
 CC -1- SIMILARITY: BELONGS TO THE ADENOVIRUS E4 30 TO 34 KD PROTEIN
 CC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC DR EMBL: M37187; AAA42514.1;
 CC DR PIR: A38519; 04ADM1.
 CC NM Early protein.
 CC SO SEQUENCE 289 AA; 33016 MW; F23D5C814DC63823 CRC64;

Query Match 11.7%; Score 186.5; DB 1; Length 289;
 Best Local Similarity 24.8%; Pred. No. 5.0e-11;
 Matches 68; Conservative 43; Mismatches 120; Indels 43; Gaps 11;

QY 52 NLTLMHNVSYRGPCSYGFTLQEWV-----VPMQVLTREELVILRK 95
 DB 20 NPLTMS--ATLSRDCCVGLPHCMGWLPSYGAVALFVMEGALPIPMXSYLSYDKHVLQ 77
 QY 96 CMHWCLCCANIDIMTSM--MIGYESMALHCHCSPGS-----LQCIAGGVYLAEMFR 146
 DB 78 L--DCLGFLPYCYQSFSTSYTGDEVMSLHCHGROGSLQVPRRLQCLAAKARELYVQ 135
 QY 147 MYVDGAMNORFIWREVVNTNMPREVMSSVFMGRHLIYRLMYDGHVSVVPM-- 205
 DB 136 KELLGTRENEYPOYRHNANKRYVNGLEYSVWC--GHFIYVA--DGAFAFLKGRRA 192
 QY 205 SPFGSALHC--GILNNIVVLCSCYCA--DSEIRVCCARRTRRLRAVRIIAETETAM-- 261
 DB 193 RLGQGLTFEESLSCVYVIVCQACAPPTDQVDHCHMLSTFLRMQNLILGRGSSPL 252
 QY 261 -----LYSCTERRRQCFRALLQHRPIIMHDY 289
 DB 253 IPGFIDPRNRTERLRQMLRHFREYKTYPIRLTY 286

RESULT 10
 UROM_BOVIN STANDARD: PRT: 643 AA.
 ID UROM_BOVIN
 AC P48733;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE UROMODULIN PRECURSOR (TAMM-HORSFALL URINARY GLYCOPROTEIN) (THP).
 GN UROMOD.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
 OC Bovinae; Bos.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY;
 RX MEDLINE: 95143938.
 RA Yu H., Papa F., Sukhatme V.P.;
 RT "Bovine and rodent tamm-horsfall protein (THP) genes: cloning,
 RT structural analysis, and promoter identification."
 RL Gene Expr. 4:63-75(1994).
 CC -1- FUNCTION: NOT KNOWN. MAY PLAY A ROLE IN REGULATING THE CIRCULATING
 CC ACTIVITY OF CYTOKINES AS IT BINDS TO IL-1, IL-2 AND TNF WITH HIGH
 CC AFFINITY (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR,
 CC THEN CLEAVED TO PRODUCE A SOLUBLE FORM WHICH IS SECRETED IN
 CC URINE (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 ZP DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: S75958; AAB3312.1;
 CC DR HSSP: P35555; IEMO.
 CC DR PRINTS: PR00023; ZPELUCIDA.
 CC DR PROSITE: PS00682; ZP_DOMAIN; 1.
 CC DR PROSITE: PS00010; ASX_HYDROXYL; 2.
 CC DR PROSITE: PS00022; EGF_1; FALSE_NEG.
 CC DR PROSITE: PS01186; EGF_2; 3.
 CC DR PROSITE: PS01187; EGF_CA; 2.
 CC DR PFAM: PF00008; EGF; 3.
 CC DR PFAM: PF00100; zona_dellucida; 1.
 CC DR KW Glycoprotein; Signal; Membrane; GPI-anchor; EGF-like domain.
 CC FT SIGNAL 1 26
 CC FT CHAIN 27 643
 CC FT DOMAIN 30 66
 CC FT DOMAIN 67 109
 CC FT DOMAIN 110 151
 CC FT DOMAIN 336 587
 CC FT DISULFID 34 43
 CC FT DISULFID 37 52
 CC FT DISULFID 54 65
 CC FT DISULFID 71 85
 CC FT DISULFID 79 94
 CC FT DISULFID 96 108
 CC FT DISULFID 114 128
 CC FT DISULFID 122 137
 CC FT DISULFID 139 150
 CC FT CARBOHYD 40 40
 CC FT CARBOHYD 78 78
 CC FT CARBOHYD 134 134
 CC FT CARBOHYD 234 234
 CC FT CARBOHYD 277 277
 CC FT CARBOHYD 324 324
 CC FT CARBOHYD 398 398
 CC FT CARBOHYD 449 449
 CC FT CARBOHYD 515 515
 CC SO SEQUENCE 643 AA; 69898 MW; 448984AB01DFA477 CRC64;

Query Match 5.7%; Score 91.5; DB 1; Length 643;
 Best Local Similarity 22.4%; Pred. No. 0.26;
 Matches 53; Conservative 19; Mismatches 86; Indels 79; Gaps 13;

QY 48 LPECNLTLMHNVSYRG-----LPCSYGFTL-----IOEMVPMQVLTREEL 90
 DB 68 LDECAVLGAHNCSTFKSCVNTLGSITVCCEGLLSLSELCEDVDECAERG----- 119
 QY 91 VILKRCMHVCLCCANIDIMTSMHIGYESMALHCHCSPGS-----LQCIAGGVYLAEMFR 146
 DB 119 --LSRCHALATCINGSENYSCVAGYLGGRHCEC--SPSCPGGLDCAVEGDAL----- 171
 QY 147 MYVDGAMNORFIWREVVNTNMPREVMSSVFMGRHLIYRLMYDGHVSVVPM-- 204
 DB 171 VCPBPCQVHRILDEYKSTEGSG-----YICDVSILG-----WRYFGQAGVRLPET 218
 QY 205 SPFGSALHC-----GILNNIVVLCSCYCA--DSEIRVCCARRTRRLRAVRIIAETETAM-- 261
 DB 219 C--VPELHCNTAPRWMLNGTTPSSDEGLVNVACAHMSGGDCCLMDA---PIQYKACA 270

RESULT 11
 EAR_EBV STANDARD: PRT: 191 AA.
 ID EAR_EBV
 AC P03182;
 DT 21-JUL-1986 (Rel. 01, Created)

[illegible]

FT DOMAIN 99 113 EXTRACELLULAR.
 FT TRANSMEM 114 133 3 (POTENTIAL).
 FT DOMAIN 134 152 CYTOPLASMIC.
 FT TRANSMEM 153 176 4 (POTENTIAL).
 FT DOMAIN 177 202 5 (POTENTIAL).
 FT TRANSMEM 203 230 5 (POTENTIAL).
 FT DOMAIN 231 252 6 (POTENTIAL).
 FT TRANSMEM 253 276 6 (POTENTIAL).
 FT DOMAIN 277 284 EXTRACELLULAR.
 FT TRANSMEM 285 309 7 (POTENTIAL).
 FT DOMAIN 310 334 CYTOPLASMIC.
 FT BLINDING 322 322 RETINAL CHROMOPHORE.
 FT LIPID 323 323 PALMITATE (BY SIMILARITY).
 FT LIPID 323 323 PALMITATE (BY SIMILARITY).
 FT DISULFID 110 187 BY SIMILARITY.
 FT CONFLICT 107 107 P -> Q (IN REF. 2).
 FT CONFLICT 137 137 I -> M (IN REF. 2).
 FT CONFLICT 241 241 L -> A (IN REF. 2).
 SQ SEQUENCE 354 AA; 39786 MW; CD18F49EC63DBFE CRC64;

Query Match 5.18; Score 81.5; DB 1; Length 354;
 Best Local Similarity 20.0%; Pred. No. 1.2;
 Matches 61; Conservative 33; Mismatches 94; Indels 117; Gaps 13;
 5 GVPFG-WTLRPTSRSLRTPYSDRLPEFETETRAITLEDHPLPPECNTLTHNV----- 60
 51 GLPIFMFTLFTIQHKRLRTPLNTLL-----NLVFNHFVLCGFVTMTSMHGY 102
 60 -----SYVGLGPCSVFTLIQENVVPMVAVLTREELVILRKCHVCLCCANIDIMTSM 112
 103 FIFGPTGCTIEF-----FATLGEVALMSLVLAVERIV-----VCKPMANRFGENH 152
 DB 103 FIFGPTGCTIEF-----FATLGEVALMSLVLAVERIV-----VCKPMANRFGENH 152
 QY 113 MHGYE-SWALHCHGSSPSLOCINAGOVLAWSFRMVDG-----AMENQR 157
 DB 153 AIMGAFITVIMLSCAP-----PLFGMSRIPEGMQSCGVDTYTLKEVNNES 202
 QY 158 FTWYREVYNYNPKREVMSSVFMRGRHL----- 187
 DB 203 FLYIMEVIFHTPLIVIF-----FCYGRLLCTYKEAQQOESLTOKAEKVTIRVAVIN 258
 QY 187 -----ITLRMYDGHVS-----VYPAWSFGYSALHCG-----ILNN-----I 219
 DB 259 VFLLGWVYAYAVAFITFHGSGNFGPVMTVPAPFAKSSALINPVYIVILNKQCRNCLI 318
 QY 220 VWLCC 224
 DB 319 TILCC 323

RESULT 13
 OPSD_LIMBE STANDARD: PRT; 289 AA.
 AC 042427;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE RHODOPSIN (FRAGMENT).
 GN RHO.
 OS Limnocottus bergianus.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
 OC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorphi;
 OC Scorpaeniformes; Cottoidae; Abyssocottidae; Limnocottus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE; 98086781.
 RA Hunt D.M., Fitzgibbon J., Slobodanyuk S.J., Bowmaker J.K.,
 RA Dulai K.S.;
 RA "Molecular evolution of the cottoid fish endemic to Lake Balkal
 RA deduced from nuclear DNA evidence".
 RL Mol. Phylogenet. Evol. 8:415-422(1997).
 CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
 CC MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY

CC LINKED TO CIS-RETINAL.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
 CC VISION IN DIM LIGHT.
 CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
 CC BE PHOSPHORYLATED (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC OPSIN SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U97270; AAB61724.1; -.
 CC GCRDB; GCR_1264; -.
 CC PROSITE; PS00237; G-PROTEIN-RECEPTOR; 1.
 CC PROSITE; PS00238; OPSIN; 1.
 CC PFM; PFM0001; 7tm.1; 1.
 CC DR Phosphorylation; Retinal protein; Transmembrane; Glycoprotein; Vision;
 CC DR Photoreceptor; Lipoprotein; Palmitate; G-protein coupled receptor.
 CC KW NON_TER 1
 CC FT DOMAIN <1 7
 CC FT TRANSMEM 8 32 EXTRACELLULAR (POTENTIAL).
 CC FT DOMAIN 33 44 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 45 69 2 (POTENTIAL).
 CC FT DOMAIN 70 84 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 85 104 3 (POTENTIAL).
 CC FT DOMAIN 105 123 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 124 147 4 (POTENTIAL).
 CC FT DOMAIN 148 173 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 174 201 5 (POTENTIAL).
 CC FT DOMAIN 202 223 CYTOPLASMIC (POTENTIAL).
 CC FT TRANSMEM 224 247 6 (POTENTIAL).
 CC FT DOMAIN 248 255 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 256 280 7 (POTENTIAL).
 CC FT DOMAIN 281 >289 CYTOPLASMIC (POTENTIAL).
 CC FT DISULFID 81 158 BY SIMILARITY.
 CC FT BLINDING 267 267 RETINAL CHROMOPHORE (BY SIMILARITY).
 CC FT CARBOHYD 171 171 POTENTIAL.
 CC FT NON_TER 289
 CC SQ SEQUENCE 289 AA; 32445 MW; 1B381201ACB80A2 CRC64;

Query Match 5.0%; Score 79; DB 1; Length 289;
 Best Local Similarity 24.0%; Pred. No. 1.7;
 Matches 36; Conservative 22; Mismatches 54; Indels 38; Gaps 8;
 55 TMHNSYVRLGICSV-GF--TLIQENVVPMVAVLTREELVILRKCHVCLCCANIDIMTS 111
 DB 69 SMHGYSVLGRGILLEGFFATLGE--IALMSLV-----VALERWYVCKPISNFEPTD 122
 QY 112 MHGYE-SWALHCHGSSPSLOCINAGOVLAWSFRMVDG-----AMFNQ 156
 DB 123 HAINGLAFSWVALACAP-----PLVGMRSRIPEGMQSCGVDTYTRAEGFNN 172
 QY 157 FTWYREVYNYNPKREVMSSVFMRGRHL----- 186
 DB 173 SFVIMFVHRLPLSVIF-----FCYGRLL 198
 RESULT 14
 OPSD_POMMI STANDARD: PRT; 352 AA.
 ID OPSD_POMMI
 AC P35403;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE RHODOPSIN.
 GN RHO.

OS Pomatoschistus minutus (Sand goby).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
 CC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
 CC Perciformes; Gobioidae; Gobiidae; Pomatoschistus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=RETINA;
 RA MEDLINE; 92396703.
 RX Archer S.N., Lythgoe J.N., Hall L.;
 RT "Rod opsin cDNA sequence from the sand goby (Pomatoschistus minutus)
 RL compared with those of other vertebrates";
 RL Proc. R. Soc. Lond. B. Biol. Sci. 246:19-25(1992).
 CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
 CC MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
 CC LINKED TO CIS-RETINAL.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
 CC VISION IN DIM LIGHT.
 CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
 CC BE PHOSPHORYLATED.
 CC -1- MISCELLANEOUS: THIS RHODOPSIN HAS AN ABSORPTION MAXIMA AT 501 NM.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC OPSIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X62405; CAA44275.1; -
 DR PIR; S40688; S40688.
 DR GCRDB; GCR_0503; -
 DR PRINTS; PR00237; GPCRHHODOPSN.
 DR PRINTS; PR00238; OPSIN.
 DR PRINTS; PR00579; RHODOPSIN.
 DR PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
 DR PROSITE; PS00238; OPSIN; 1.
 DR PFAM; PF00001; 7tm1; 1.
 KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
 FT Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor;
 FT DOMAIN 1
 FT TRANSSEM 37 61
 FT TRANSSEM 62 73
 FT TRANSSEM 74 98
 FT TRANSSEM 99 113
 FT TRANSSEM 114 133
 FT TRANSSEM 134 152
 FT TRANSSEM 153 176
 FT TRANSSEM 177 202
 FT TRANSSEM 203 230
 FT TRANSSEM 231 252
 FT TRANSSEM 253 276
 FT TRANSSEM 277 284
 FT TRANSSEM 285 309
 FT TRANSSEM 310 352
 FT CARBOHYD 2
 FT BINDING 15 15
 FT LIPID 296 296
 FT LIPID 322 322
 FT LIPID 323 323
 FT DISULFID 110 187
 SQ SSOURCE 352 AA; 39409 MW; 888CE379A9ED656 CRC64;

Query Match 5.0%; Score 79; DB 1; Length 352;
 Best Local Similarity 19.9%; Pred. No. 2.1;
 Matches 51; Conservative 31; Mismatches 76; Indels 98; Gaps 13;

55 THANVSRYRLPCSV-GF-TLIGENVYPMQVLTRELYILRKCMHYCLCANIDIMTS 111

Db 98 SMHGFEVLGRGCVESGFATLIGE -IALMSLVLAVERMV-----YCKPISNRFETEN 151
 QY 112 MINGYE-SWALHCHSPSPSLGCIAGGYLASFPMVVDG-----AMFNQ 156
 Db 152 HAINGVAFSWIMATCAVP-----PLVGWSKTYPEGMQSCGVYYTRAGFENNE 201
 QY 157 REIWEREVNVNMPKEVMSVFMGRHLYLR----- 191
 Db 202 SEVLYMEIVHFLAPLYIF-----FCYGRLLCAVKKAAQAQSEETTORAREEYTRAVTM 257
 QY 191 -----LWY-----DGHGVSVPMSGYSALHCGILNINIVLCCS--- 226
 Db 258 VIGFLSWLPYASVAVYITHOCTEGRPLEMTPAFKSSALY-----NPMIYCMKNQF 313
 QY 226 -YCADISEIRVRCAR 240
 Db 314 RHCM-----ITTLCCGR 325

RESULT 15
 OPED_ZEUFA ID OPED_ZEUFA STANDARD; PRT; 354 AA.
 AC 042604;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 38, Last annotation update)
 DE RHODOPSIN.
 GN RHO.
 OS Zeus faber ('John Dory').
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
 CC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;
 CC Perciformes; Zeidae; Zeus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RA Archer S.N., Hirano J.;
 RT Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 RL -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
 CC MEDIANE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
 CC LINKED TO CIS-RETINAL.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: ROD SHAPED PHOTORECEPTOR CELLS WHICH MEDIATES
 CC VISION IN DIM LIGHT.
 CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
 CC BE PHOSPHORYLATED.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC OPSIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Y14484; CAA74832.1; -
 DR GCRDB; GCR_2445; -
 DR PRINTS; PR00237; GPCRHHODOPSN.
 DR PRINTS; PR00238; OPSIN.
 DR PRINTS; PR00579; RHODOPSIN.
 DR PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
 DR PROSITE; PS00238; OPSIN; 1.
 DR PFAM; PF00001; 7tm1; 1.
 KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
 FT Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor;
 FT DOMAIN 1
 FT TRANSSEM 37 61
 FT TRANSSEM 62 73
 FT TRANSSEM 74 98
 FT TRANSSEM 99 113
 FT TRANSSEM 114 133
 FT TRANSSEM 134 152
 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 153 178 4 (POTENTIAL).
 FT DOMAIN 177 202 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 203 230 5 (POTENTIAL).
 FT DOMAIN 231 252 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 253 276 6 (POTENTIAL).
 FT DOMAIN 277 284 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 285 309 7 (POTENTIAL).
 FT DOMAIN 310 354 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 2 2 BY SIMILARITY.
 FT CARBOHYD 15 15 BY SIMILARITY.
 FT DISULFID 110 187 RETINAL CHROMOPHORE (BY SIMILARITY).
 FT BINDING 296 296 PALMITATE (BY SIMILARITY).
 FT LIPID 322 322 PALMITATE (BY SIMILARITY).
 FT LIPID 323 323 PALMITATE (BY SIMILARITY).
 SQ SEQUENCE 354 AA; 39662 MW; A97D38F8E86DB3 CRC64;

Query Match 5.08; Score 79; DB 1; Length 354;
 Best Local Similarity 18.78; Pred. No. 2.1;
 Matches 53; Conservative 35; Mismatches 87; Indels 108; Gaps 14;

OY 55 TWNNVSYVGLPCSY-GE--TLIOEYVVPMDVLTREELVLRKCMHYCLCCANIDIMTS 111
 DB 98 SMHGFYVIGRTGCTNCEGFATLGE-IALMSIVYLAVRMVY-----VCKPISNFRGEN 151
 OY 112 KMIRGE-SMALHCHCSSPGSLQCIAGQVLAWEFMVVDGAMF-----NQ 156
 DB 152 HAVMGVSFTWLACACVP-----PLFGMSRYIPEGMQCSGIDYTRAPGVNNE 201
 OY 157 REIYREVYNNMPKEVEMSSVFMRGRHLI----- 188
 DB 202 SPVITMEVCHFSIPLTIIP---FCYGRLLCAVKDAAAQOSEETTORAREVSRMVIM 257
 OY 188 ---YRLWTD-----GHGVSVPAMSFYSALHCGILNNIVLCCS--- 226
 DB 258 VIGFLICMLPYASVAMFIFTHOGSEFGPEFTIPA---FFAKSSAIYNPMIYICMKNOP 313
 OY 226 -YCADLSEIRVRCARTRRLMLRAVRIAEETMLYSCRT 267
 DB 314 RHCM---ITTLCCGKNP-----FEEEGASTASKTE 342

Search completed: June 23, 2000, 10:14:48
 Job time: 669 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2000, 09:51:57 ; Search time 61.44 "Seconds

(without alignments)
331.775 Million cell updates/sec

Title: US-09-214-478-2

Perfect score: 1594
Sequence: 1 MTSGVPEGMKLRPRSRSL.....RALLQHRPILMHVDSTPM 294

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

SPTREMBL.12.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_ricent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	223.5	14.0	219	071106	071106 bovine aden
2	180	11.3	163	084211	084211 porcine ade
3	180	11.3	163	09YTR0	09YTR0 porcine ade
4	163.5	10.3	268	12 071108	071108 bovine aden
5	108	6.8	220	P89027	P89027 ovine adeno
6	90	5.6	1293	09Y748	09Y748 emericella
7	87	5.5	494	044515	044515 caenorhabd1
8	85.5	5.4	211	067148	067148 aquifex neo
9	85.5	5.4	954	0597180	0597180 drosophila
10	84.5	5.3	431	023017	023017 arabidopsis
11	83	5.2	354	13 09W6A8	09W6A8 brachydanio
12	82.5	5.2	181	10 023665	023665 arabidopsis
13	82.5	5.2	405	10 024416	024416 fragaria an
14	82	5.1	1603	021884	021884 caenorhabd1
15	80.5	5.1	433	09Y285	09Y285 plasmodium
16	80	5.0	353	13 09YH02	09YH02 sparus aua
17	79.5	5.0	456	022130	022130 caenorhabd1
18	79.5	5.0	1259	6 018735	018735 canis fam11
19	79	5.0	353	13 09YH05	09YH05 diplopus an
20	79	5.0	353	13 09YH03	09YH03 sarpa salpa

21	78	4.9	353	13 09YH04	09YH04 diplopus vu
22	77.5	4.9	459	8 092258	092258 canis fam11
23	77.5	4.9	530	12 09YMH3	09YMH3 lymantria d
24	77.5	4.9	642	2 0928V7	0928V7 streptomyce
25	77.5	4.9	3008	12 039929	039929 hepatitis c
26	77	4.8	227	10 023668	023668 arabidopsis
27	77	4.8	353	13 09YH00	09YH00 lithognathu
28	76	4.8	505	2 059435	059435 escherichia
29	76	4.8	932	10 082748	082748 arabidopsis
30	75.5	4.7	373	5 022188	022188 caenorhabd1
31	75.5	4.7	407	12 040917	040917 kaposi's sa
32	74.5	4.7	303	5 017080	017080 caenorhabd1
33	74.5	4.7	321	3 008959	008959 saccharomyc
34	74.5	4.7	407	12 088909	088909 kaposi's sa
35	74.5	4.7	895	5 045631	045631 caenorhabd1
36	74	4.6	353	13 09Y628	09Y628 chelon labr
37	74	4.6	3072	12 092645	092645 clover yell
38	74	4.6	295	12 011425	011425 avian adeno
39	73.5	4.6	315	4 050403	050403 homo sapien
40	73.5	4.6	346	8 021195	021195 megachasma
41	73.5	4.6	348	11 09WTK1	09WTK1 cavia porce
42	73.5	4.6	353	13 09Y629	09Y629 mugil cepha
43	73	4.6	353	13 09Y626	09Y626 liza aurata
44	73	4.6	367	5 P91382	P91382 caenorhabd1
45	73	4.6			

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	219 AA.
071106	071106			
AC	071106;			
DT	01-AUG-1998 (T-EMBLrel. 07, Last sequence update)			
DT	01-AUG-1998 (T-EMBLrel. 07, Last sequence update)			
DT	01-NOV-1998 (T-EMBLrel. 08, Last annotation update)			
DE	ADENOVIRUS 3 COMPLETE GENOME.			
OS	Bovine adenovirus type 3 (Mastadenovirus bovis).			
OC	Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-MBR-1;			
RA	LEE J.B., BABIUK M.K., IDAMAKANTI N., REDDY P.S., ZAKHARTCHOUK A.N.,			
RA	PINE C., BABIUK L.A., TIKOO S.K.;			
RL	Viruses Genes 0:0-0(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-MBR-1;			
RA	BABIUK M.K., REDDY P.S., ZAKHARTCHOUK A.N., IDAMAKANTI N., PINE C.,			
RA	BABIUK L.A., TIKOO S.K.;			
RL	Viruses Genes 0:0-0(1998).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-MBR-1;			
RA	MEDLINE: 98105785.			
RA	REDDY P.S., IDAMAKANTI N., ZAKHARTCHOUK A.N., BAXI M.K., LEE J.B.,			
RA	PINE C., BABIUK L.A., TIKOO S.K.;			
RT	"Nucleotide sequence, genome organization, and transcription map of			
RT	bovine adenovirus type 3."			
RL	J. Virol. 72:1394-1402(1998).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-MBR-1;			
RA	REDDY P.S., IDAMAKANTI N., ZAKHARTCHOUK A.N., BAXI M.K., LEE J.B.,			
RA	PINE C., BABIUK L.A., TIKOO S.K.;			
RL	Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF030154; AAD09737.1;			
SO	SEQUENCE 219 AA; 26028 MW; 092D5CCD CRC32;			

Query Match 14.0%; Score 223.5; DB 12; Length 219;
Best Local Similarity 23.3%; Pred. No. 7.7e-16;

Matches 56; Conservative 42; Mismatches 106; Indels 17; Gaps 5;

QY 77 WVPMDVLTREELVIRKCHV--CLC--CANIDT--WTSMAHGYESWALHCHCSP 129
 Db 4 WPIRMKVIIMPED-----AHLTSCCDYCLSPVIGKRKRSYVGEELTYHCHCPK 56
 QY 130 GSLOCAGGGOVLASWFMVVDGAMFNORFTMYREVYNNKFKVMSSVFMGRHLTYL 189
 Db 57 RSLQCRSAAYVFGVLVACQVHGAQFDRVYVYFRELSTKSLSVLAVGSVYNNVHLTYF 116
 QY 190 RLVWDGVSVVPMASGYSALHCGILNNIVYLCSTCADLSEIRVCCARTRLMRA 249
 Db 117 NSTFMIEGSIYRCTPGECLWKGNSCYLLILICRSCNPLEMSVNCCLKCKOKIRM 176
 QY 250 VRIAEETAMLXSCRTERRROOFIRALDHOHPITLMDHD 290
 Db 177 LNLAPKRVKHL---GLARLRDRFLRLARFREPVLFDKID 214

RESULT 2
 Q84211
 ID 084211 PRELIMINARY; PRT; 163 AA.
 AC 084211;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-NOV-1996 (TREMBLrel. 01, Last annotation update)
 DE MASTADENOVIRUS 19.8 KDA POLYPEPTIDE (ORF 8).
 OS Porcine adenovirus 3.
 NC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PAV-3;
 RA REDDY S., IDAMAKANTI N., DERBYSHIRE J.B., NAGY E.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: L43363; CAB24462.1;
 KW Repeat.
 SQ SEQUENCE 163 AA; 19786 MW; DAE89A1D CRC32;

Query Match
 Best Local Similarity 11.3%; Score 180; DB 12; Length 163;
 Matches 45; Conservative 20; Mismatches 49; Indels 28; Gaps 4;

QY 141 LASFRMYVDGAMFNORFTMYREVYNNKFKVMSSVFMGRHLTYLMD----- 195
 Db 1 MRSIMREIIRCEYRIRTMWFRVAVNLPVSRYVGVSGVFMGRHLMYIKIMYDDCKL 60
 QY 195 --GHVGSVPMASGYSALHC-GILNNIVYLCSTCADLSEIRVCCARTRLMRAVR 251
 Db 61 RMFMGEVI-----YCGDGYANYIVLCRCHELSEPVARRCAQRLRMWKLAAE 110
 QY 252 IIAEETAMLXSCRTERRROOF 273
 Db 111 ALGAORRG-----RKQOF 123

RESULT 3
 Q9YTR0
 ID 09YTR0 PRELIMINARY; PRT; 163 AA.
 AC 09YTR0;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
 DE 19.8 KDA POLYPEPTIDE.
 OS Porcine adenovirus 3.
 NC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-6618;
 RA REDDY P.S., NAGY E., DERBYSHIRE J.B.;
 RT "Sequence analysis of putative pVIIT, E3 and fibre regions of porcine adenovirus type 3."
 RT

RL Virus Res. 36:97-106(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-6618;
 RX MEDLINE; 95407103.
 RA REDDY P.S., TUDOLY T., DENNIS J.R., DERBYSHIRE J.B., NAGY E.;
 RT "Comparison of the inverted terminal repetition sequences from five porcine adenovirus serotypes."
 RL Virology 212:237-239(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-6618;
 RX MEDLINE; 98014367.
 RA REDDY P.S., IDAMAKANTI N., DERBYSHIRE J.B., NAGY E.;
 RT "Characterization of the early region 4 of porcine adenovirus type 3."
 RL Virus Genes 15:87-90(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-6618;
 RX MEDLINE; 99058191.
 RA REDDY P.S., IDAMAKANTI N., SONG J.Y., LEE J.B., HYUN B.H., PARK J.H.,
 RA CHA S.H., BAE Y.T., TIKOO S.K., BABIUK L.A.;
 RT "Nucleotide sequence and transcription map of porcine adenovirus type 3."
 RL Virology 251:414-426(1998).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-6618;
 RA REDDY P.S.;
 RT "Sequence and transcription map analysis of early region-1 of porcine adenovirus type 3."
 RL Virus Res. 0:0-0(1999).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN-6618;
 RA REDDY P.S.;
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF083132; AAC99443.1;
 KW SEQUENCE 163 AA; 19786 MW; DAE89A1D CRC32;

Query Match
 Best Local Similarity 11.3%; Score 180; DB 12; Length 163;
 Matches 45; Conservative 20; Mismatches 49; Indels 28; Gaps 4;

QY 141 LASFRMYVDGAMFNORFTMYREVYNNKFKVMSSVFMGRHLTYLMD----- 195
 Db 1 MRSIMREIIRCEYRIRTMWFRVAVNLPVSRYVGVSGVFMGRHLMYIKIMYDDCKL 60
 QY 195 --GHVGSVPMASGYSALHC-GILNNIVYLCSTCADLSEIRVCCARTRLMRAVR 251
 Db 61 RMFMGEVI-----YCGDGYANYIVLCRCHELSEPVARRCAQRLRMWKLAAE 110
 QY 252 IIAEETAMLXSCRTERRROOF 273
 Db 111 ALGAORRG-----RKQOF 123

RESULT 4
 Q71108
 ID 071108 PRELIMINARY; PRT; 268 AA.
 AC 071108;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE ADENOVIRUS 3 COMPLETE GENOME.
 OS Bovine adenovirus type 3 (Mastadenovirus bos3).
 NC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WB-1;
 RA LEE J.B., BAXI M.K., IDAMAKANTI N., REDDY P.S., ZAKHARCHOUK A.N.,

RA PYNE C., BABIUK L.A., TIKOO S.K.;
 RL Virus Genes 0:0-0(1998).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WB-1;
 RA BAXI M.K., REDDY P.S., ZAKHARCHOUK A.N., IDAMAKANTI N., PYNE C.,
 BABIUK L.A., TIKOO S.K.;
 RL Virus Genes 0:0-0(1998).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WB-1;
 RA MEDLINE: 98105765.
 RA REDDY P.S., IDAMAKANTI N., ZAKHARCHOUK A.N., BAXI M.K., LEE J.B.,
 PYNE C., BABIUK L.A., TIKOO S.K.;
 RL "Nucleotide sequence, genome organization, and transcription map of
 RT bovine adenovirus type 3";
 RL J. Virol. 72:1394-1402(1998).
 [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-WB-1;
 RA REDDY P.S., IDAMAKANTI N., ZAKHARCHOUK A.N., BAXI M.K., LEE J.B.,
 PYNE C., BABIUK L.A., TIKOO S.K.;
 RL Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF030154; AAD09739.1;
 SQ SEQUENCE 268 AA; 31016 MW; BEA16AF CRC32;
 Query Match 10.3%; Score 163.5; DB 12; Length 268;
 Best Local Similarity 22.0%; Pred. No. 2e-09; Mismatches 92; Indels 17; Gaps 6;
 Matches 44; Conservative 47;
 QY 100 CACGACNIDIMTSMHIGESMALHCHCSSPG--SLQCIAGGVLAWEFVVDGAMFNOR 157
 DB 59 CDDCTGRCDDSEFFISGDATWRYC-CSKQADOSLOGLCRATILMDLFMCVAGAFDPM 117
 QY 158 FIVREVVYNNMPEKVENSVFMGRHLIYRLMTDGHVS-----VVPMSFGYSAL 211
 DB 118 FPFRELSLSTRYIYVGSVYNNRVHLLY---FTNVGASNOLKSLPFANGNYM 173
 QY 212 HCGILNNIVVLCSCYCADSEIRVRCARTRMLRAVRIIAETMTALYSCRTERRRO 271
 DB 174 GSGLDKCIIPCYDAELSEVSASRCLLVKEMIMCYITCLPSDR--YQCGLEPHND 230
 QY 272 QFIRALLQHRPIIMHDYDS 291
 DB 231 KLIQRYAA-GEPIVKRYFDR 249
 RESULT 5
 P89027 PRELIMINARY; PRT; 220 AA.
 AC P89027;
 DT 01-MAY-1997 (Tremblrel. 03, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
 DE E4.2 PROTEIN (25.6 KD PROTEIN).
 OS Ovine adenovirus.
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Atadenovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OAV287;
 RA MEDLINE: 96240641.
 RA YRATI S., BROOKES D.E., STRIKE P., KHATRI A., BOYLE D.B., BOTH G.W.;
 RT "Unique genome arrangement of an ovine adenovirus: identification of
 RT new proteins and proteinase cleavage sites";
 RL Virology 220:186-199(1996).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OAV287;
 RA MEDLINE: 97271311.
 RA XU Z.Z., HYATT A., BOYLE D.B., BOTH G.W.;
 RT "Construction of ovine adenovirus recombinants by gene insertion or
 RT deletion of related terminal region sequences.";

RL Virology 230:62-71(1997).
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OAV287;
 RX MEDLINE: 96277079.
 RA KHATRI A., BOTH G.W.;
 RT "Identification of transcripts and promoter regions of ovine
 RT adenovirus OAV287";
 RL Virology 245:128-141(1998).
 [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OAV287;
 RA BOTH G.W.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U40839; AAD45954.1;
 SQ SEQUENCE 220 AA; 25401 MW; DCAF782 CRC32;
 Query Match 6.8%; Score 108; DB 12; Length 220;
 Best Local Similarity 27.7%; Pred. No. 0.0011;
 Matches 43; Conservative 22; Mismatches 50; Indels 40; Gaps 10;
 QY 123 HCHGSSPSGLQCIAGGVLAWEFVVDGAMFNORFIWREY-VVNNMPEKVENSVF 180
 DB 78 HCHGSSPSGLQCIAGGVLAWEFVVDGAMFNORFIWREY-VVNNMPEKVENSVF 135
 QY 181 MGRHLIY-----RLMTDGHVGSVVPMSFGYSALHCGILNNIVVLCSCYCADL 230
 DB 136 MGRHLIY-----RLMTDGHVGSVVPMSFGYSALHCGILNNIVVLCSCYCADL 184
 QY 231 SEIRVRCARTRMLRAVRIIAETMTALYSC 264
 DB 185 LQ---QCISREKKNFLR-----MLYCC 205
 RESULT 6
 Q9Y748 PRELIMINARY; PRT; 1293 AA.
 ID Q9Y748;
 AC Q9Y748;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
 DE ABC TRANSPORTER PROTEIN ATRC.
 GN ATRC.
 OS Emericella nidulans (Aspergillus nidulans).
 OC Eukaryota; Fungi; Ascomycota; Euscomycetes; Plecomycetes;
 CC Eurotiales; Trichocomaceae; Emericella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-FGSC A4;
 RX MEDLINE: 99156773.
 RA ANGERMAYR K., PARSON W., STOPFLER G., HAAS H.;
 RT "Expression of atrc - encoding a novel member of the ATP binding
 RT cassette transporter family in Aspergillus nidulans - is sensitive to
 RT cycloheximide";
 RL Biochim. Biophys. Acta 1453:304-310(1999).
 DR EMBL: AF082072; AAD25925.1;
 DR PROSITE: PS00211; ABC_TRANSPORTER; 2.
 KW ATP-binding; Transport.
 SQ SEQUENCE 1293 AA; 141984 MW; 2DBE0446 CRC32;
 Query Match 5.6%; Score 90; DB 3; Length 1293;
 Best Local Similarity 21.4%; Pred. No. 0.65;
 Matches 50; Conservative 36; Mismatches 94; Indels 54; Gaps 10;
 QY 90 IYLKRCMHVCLCAN-----IDMTSMHIGESMALHCHCSSPSGLQCI 136
 DB 717 LIVLYRCRQ-CLPSSGRQSHLIFQPLPAMASEMLHKGFMGIIVYLAVALYCYA 775
 QY 137 G-----GOVLASFRAVVDGAMFNORFIWREY-----VVNNM---PREVFMSSY 179
 DB 776 GIGFFFTVAASFLSGYRSRYRAAMLNDVSFEEDDSAGVMTQSLTDIPORIEDLISL 835

QY 180 FMGRHLLYLRLMTYDGVYVPMSPGYSALHGLINNIYVLCSCADLSEIRVR-CC 238
 DB 836 CLGFLIVVAV-----JASCIILAVGWRLALVAFGCPPL---FLAYVVRLEITC 887
 QY 239 ARFRRLMLR-----AVRIIA-----EETAMLSGRTERRRQOFTALL 278
 DB 888 QERTRLYESARFATEAISAFITVASLLEEKVIOMYDBRLSHTSPKFRIRIL 941

RESULT 7

AC 044515 PRELIMINARY; PRT; 494 AA.
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE F56A11.4 PROTEIN.
 GN F56A11.4
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 CC Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE, 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BORTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HARKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATRILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERRY C., KIRKEN L., ROOPER A., SANDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., STURTON J.,
 RA THIERRY-MIGES J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA GATUNG S., GOELA D., HARPER M.;
 RL Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL, AF038619; AAB92076.1; F59CC020 CRC32;
 SQ SEQUENCE 494 AA; 56780 MW; F59CC020 CRC32;

Query Match 5.5%; Score 87; DB 5; Length 494;
 Best Local Similarity 20.2%; Pred. No. 0.45;
 Matches 51; Conservative 40; Mismatches 96; Indels 66; Gaps 10;

QY 36 ETRATILDEHPLLEPCNLTLMANVSYVGLPCSVGFTLIQGVVPMVMTREELVILRK 95
 DB 53 EINDTCAEDP---PADAFRMLIYITIGTVCSIGI-----VINTPLILSLRLDVRFS 103
 QY 96 CMHVCILCCANDIMTS---MMIHG-----YESMALCHCSPGSLQCIAGGQ 139
 DB 104 NIIYFLIACDILVLECFMILFASLVMDYFRELALTYCHWFYIKYVS-----TVGQ 156
 QY 140 VASAFRNVVDGAMFNORFIYREVYNIMKREVMNS---SVFMRGHLIYLRMYD 194
 DB 157 VILASTLLIYVAAEFYICSLKSSIQFSPQRFLFISYGACALFKKGSVFPELEL--- 214
 QY 195 GHVGSVPMSPGYSALHGLINNIYVLCSCADLSEIRVRCAARTRLMRAVRIIA 254
 DB 214 -----OSLPCEPFOVRL-----DLSEI-----TRTKYLIYGLKIFS 248

QY 255 EE-TTAMLYSCRT 266
 DB 249 ESKTMMFMCRS 261

RESULT 8

ID 067148 PRELIMINARY; PRT; 211 AA.
 AC 067148
 DT 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
 DE FORMATE DEHYDROGENASE GAMMA SUBUNIT.
 GN FDOI.
 OS Aquifex aeolicus.
 CC Bacteria; Aquificales; Aquificaceae; Aquifex.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VF5;
 RX MEDLINE, 98196666.
 RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
 RA GRAHAM D.E., OVERBEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus.";
 RL Nature 392:353-358(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VF5;
 RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
 RA GRAHAM D.E., OVERBEK R., SNEAD M.A., KELLER M., AUJAY M., HUBER R.,
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
 RL Submitted (JUN-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL, AB000720; AF007109.1;
 SQ SEQUENCE 211 AA; 24297 MW; 39A33D85 CRC32;

Query Match 5.4%; Score 85.5; DB 2; Length 211;
 Best Local Similarity 23.4%; Pred. No. 0.25;
 Matches 36; Conservative 20; Mismatches 35; Indels 63; Gaps 7;

QY 64 GLPCSVGFTLIQGVVPM---DMVLTREELVILRK-----M 97
 DB 67 GYVFSIGAVLM---ILKWRDVMITREDIVMLTKAVISGRHELPVYKYNAGOLIXA 123
 QY 98 HVCLOCANIDIMTSMI-----HGYSMALCHC-----CSSPGS 131
 DB 124 WVFTLSAIVFELTGLIMFEPENFSGLVMSVTHEITFIINGAFTIHIYMATVGPVS 183
 QY 132 LQCIAGGVYVPMSPGYSALHGLINNIYVLCSCADLSEIRVRCAARTRLMRAVRIIA 254
 DB 184 IMGIGKVSATV-----AKFHHP-KVYREV 209

RESULT 9

ID 097180 PRELIMINARY; PRT; 954 AA.
 AC 097180
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
 DE RASGAP PROTEIN.
 GN RASGAP.
 OS Drosophila melanogaster (Fruit fly).
 CC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OREGON-R;
 RX MEDLINE, 99147028.
 RA FELDMAN P., EICHER E.N., LEJEVENS S.J., HAFEN E., HUGHES D.A.;
 RT "Control of growth and differentiation by Drosophila RasGAP, a homolog

RT of p120 Ras-GTPase-activating protein.
 RL Mol. Cell. Biol. 19:1928-1937 (1999).
 DR EMBL: AJ012609; CAI0073.1;
 DR HSSP: P20936; 1MER.
 DR PROSITE: PS00509; RAS-GTPASE-ACRIV-1;
 SQ SEQUENCE 954 AA; 107950 MW; 0DB2B774 CRC32;

Query Match 5.48; Score 85.5; DB 5; Length 954;
 Best Local Similarity 22.38; Pred. No. 1.4;
 Matches 59; Conservative 37; Mismatches 85; Indels 83; Gaps 14;

QY 63 RGPCSGVFTLLI-----QEWVVPW-----DVLTRRELYI--LRKCMHCLCCANIDI 108
 DB 454 RALPCLATVYLKAPSGSESYEWINSKACQDSQSLSPAOKVSLRE-----LRCLN--- 506
 QY 109 MTSKMHIGYESMAL-----HCRCSPGSLQCIAGGOVLASWFRMVVDGAMFNORFIWYR 162
 DB 506 -----LHVLLEAHLRPFKLVPHPYCS-----ISLQVKYKTKYKI-----APEWEE 548
 QY 163 EVVYVNNPKYEMVMSVFM-----RGRHLIYLRL-----WYDGHVGSVVA 203
 DB 549 EFWVDVPPDVVSULTLILISGKRGKSEVAELTIDLSLKNQETEGW--QITGTPM 606
 QY 204 MSFGYSLHCGILNNIVVLCSCYCADLSEIRVRCARTRRLMDRAVRIAE----- 256
 DB 607 GEMSLRLRMRYLDDLMIPCEEY--SPLOQLLES-----ELYAVKALALCHNDRVP 657
 QY 256 ETYALISCTRRERRRROPTIRLLQ 279
 DB 658 LATALLRVROEKRETELIRMLCQ 681

RESULT 10

Q23017 PRELIMINARY; PRT; 431 AA.

AC 023017;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE TIGL1.7 PROTEIN.
 GN TIGL1.7.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eustoids II; Brassicales; Brassicaceae;
 OC Arabidopsae.

OC [1]
 RN RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA OSBORNE B.I., VYSOTSKAIA V.S., TORIOMI M., YU G., OJI O., LIU S.,
 RA LI J., HONG L., BUEHLER E., CONWAY A.B., CONWAY A.R., DEMAR K.,
 RA FENG J., KIM C., KURTZ D., LI Y., SHINN P., SUN H., DAVIS R.W.,
 RA ECKER J.R., FEDERSPIEL N.A., THEOLOGIS A.;
 RN Submitted (Aug-1997) to the EMBL/Genbank/DBJ databases.
 [2]
 RN RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA THEOLOGIS A.;
 RN Submitted (Aug-1997) to the EMBL/Genbank/DBJ databases.
 [3]
 RN RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA THEOLOGIS A.;
 RN Submitted (Aug-1997) to the EMBL/Genbank/DBJ databases.
 [4]
 RN RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA THEOLOGIS A.;
 RN Submitted (Oct-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC002376; AAB80622.1;
 DR MENDEL, 25798; Arath:1088;25798.
 DR PFAM: PF00544; pec_lyase; 1;

DR PRINTS: PRO0807; AMBALERGEN.
 SQ SEQUENCE 431 AA; 47770 MW; 868EC285 CRC32;

Query Match 5.38; Score 84.5; DB 10; Length 431;
 Best Local Similarity 21.78; Pred. No. 0.7;
 Matches 43; Conservative 35; Mismatches 73; Indels 47; Gaps 12;

QY 43 EDHPLPECNLTLMHNVSVYRGVSGFTLLQEWV--PMDVLTRELYLRKC----- 97
 DB 127 DNPVNPFRGTL-RAVIADRL-----WVFRKDVQLKQELLYNSFKTIDG 174
 QY 97 ----MHVCL--CCANIDMTSMHIGYESMALHCHCSPGSLQCIAGGOVLASWFRMVVDG 151
 DB 175 RGVNVAHANGCITIOFTVNVIVHG-----LHIHCKPTGNAMVRSSETHGCV-RTMADG 228
 QY 152 ----AMFNORFIYREVNN-----MPKRVNMSVFRGRHLI-VLRMLYDGHVS 199
 DB 229 DAISITGSSHV-----IDNSLSHCADGLVDVAKGSTATTISNNLTHHNEVMLGHSDS 284
 QY 200 VV--PAMSEFGYSLHCGI 215
 DB 285 YMRDKAMQVTLIAVNHGCV 302

RESULT 11

Q9W6A8 PRELIMINARY; PRT; 354 AA.

AC Q9W6A8;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE BLUE-SENSITIVE OPSIN.
 GN BRACHYDANIO RERIO (Zebrafish) (Zebra danio).
 OS Brachydanio rerio (Zebrafish) (Zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
 OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Cyprinidae; Rasbortinae; Danio.
 RN [1]
 RN RP SEQUENCE FROM N.A.
 RA WHITEHEAD T.S., DOHO C.J., HYDE D.R.;
 RT Cloning and characterization of six zebrafish photoreceptor opsin
 cDNAs and immunolocalization of their corresponding proteins.
 RL Vls. Neurosci. 0:0-0(1999).
 DR EMBL: AF109372; AAD24753.1;
 DR PROSITE: PS00238; OPSIN; 1.
 SQ SEQUENCE 354 AA; 39483 MW; 535F4C88 CRC32;

Query Match 5.28; Score 83; DB 13; Length 354;
 Best Local Similarity 24.18; Pred. No. 0.81; 70; Indels 38; Gaps 6;

QY 67 CSV-GFT-LIQEVVPMVMDVLTRELYLTKKCMHVCCLCCANIDMTSMHIGYESMALHC 124
 DB 117 CKIGFTISTIGAMVSLA-----VVALEKWLVTICKPLGFTFKTPHALAG----- 164
 QY 125 HCSPGSLQCIAGGOVLASWFRMVVDGAMFNORFIYREVNNVMPKRVN----- 177
 DB 164 -CIIPLFCWALAGLPPLIGMSRYIPEGLOCSGPDWYTTNNKFNNESYVMEFLCFCAVP 222
 QY 177 --SSVFRGRHLIYLRL-----WYDGHVGSVFPMSFGY 208
 DB 223 FSTIVFCYGLLITLIKLAARQADSASIQKREYVTKMAYVYVYGF 268

RESULT 12

Q23666 PRELIMINARY; PRT; 181 AA.

AC Q23666;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)

DE PUTATIVE PECTATE LYASE (FRAGMENT).
 GN A1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 Arabidopsis.
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 97422403.
 RA KULIKAVSKAS R., MCCORMICK S.;
 RT "Identification of the tobacco and Arabidopsis homologues of the
 RL pollen-expressed LAT59 gene of tomato."
 RL Plant Mol. Biol. 34:809-814(1997).
 DR EMBL: U83620; AAB69760.1; -
 DR MENDEL: 25560; Arabid:1088;25560.
 DR PFAM: PF00544; pec_lyase; 1.
 KW lyase.
 FT NON_TER
 FT NON_TER
 SQ SEQUENCE 181 AA; 20088 MW; EF3426C6 CRC32;

Query Match 5.2%; Score 82.5; DB 10; Length 181;
 Best Local Similarity 20.8%; Pred. No. 0.43;
 Matches 35; Conservative 34; Mismatches 54; Indels 45; Gaps 11;

QY 43 EDHPLPCNTLTMHNSYVGLPCSVGFTLQEMVY-PMDVLTRELYLRKC----- 97
 DB 34 DDNPNPGRPL-RHAYIQDRPL-----WIFKRMVTLQELVNSFKITDG 81
 QY 97 ---MHVCL-CCANIDITSMHIGYSWALHCHCSPGSLQCIAGQVLAFFRMVVDG 151
 DB 82 RGANYHANGCCTIOFTNTVYHG-----LHIDCKPGNAM-VRSETHFGW-RTMADG 134
 QY 152 ---AMFNORFIWREVVNYN-----MPKEVMSVFRGHLTY 188
 DB 135 DAISIFGSHW-----IDHNSLHCHADGLVRSMSGTAITTSNNLTH 178

RESULT 13
 024416 PRELIMINARY; PRT; 405 AA.
 AC 024416;
 DT 01-JAN-1998 (TRENBLREL. 05, Created)
 DT 01-JAN-1998 (TRENBLREL. 05, Last sequence update)
 DT 01-NOV-1999 (TRENBLREL. 12, Last annotation update)
 DE PEPTATE LYASE.
 GN PL.
 OS Fragaria ananassa (Strawberry).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids I; Rosales; Rosaceae; Fragaria.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-CHANDLER.
 RX MEDLINE: 97435972.
 RA MEDINA-ESCOBAR N., CARDENAS J., MOTANO E., CABALLERO J.L.,
 RA MONO-BLANCO J.;
 RT "Cloning, molecular characterization and expression pattern of a
 RT strawberry ripening-specific cDNA with sequence homology to pectate
 RT lyase from higher plants."
 RL Plant Mol. Biol. 34:867-877(1997).
 DR EMBL: U63550; AAB71208.1; -
 DR MENDEL: 26059; Arabid:1088;26059.
 DR PFAM: PF00544; pec_lyase; 1.
 KW lyase.
 FT NON_TER
 FT NON_TER
 SQ SEQUENCE 405 AA; 45744 MW; 2B9944A0 CRC32;

Query Match 5.2%; Score 82.5; DB 10; Length 405;
 Best Local Similarity 24.4%; Pred. No. 1.1;
 Matches 47; Conservative 25; Mismatches 66; Indels 55; Gaps 12;

QY 44 DHPLPCNTLTMHNSYVGLPCSVGFTLQEMVY-PMDVLTRELYLRKC----- 97
 DB 148 DDNPNPGRPL-RHAYIQDRPL-----WIFKRMVTLQELVNSFKITDG 195
 QY 97 ---MHVCL-CCANIDITSMHIGYSWALHCHCSPGSLQCIAGQVLAFFRMVVDG- 152
 DB 196 GVNHTAYGCGCTIOFTNTVYHG-----LHIDCKPGNAMVRSSPHYGM-RTMADG 249
 QY 152 ---AMFNORFIWREVVNYNMPKEVMSVFRGHLTYLRMYGHGVSVPMPSPGYS 209
 DB 250 GISIFGSHW-----VDHN-----SLSCADGLT-----DAISGTAITTSNNYF 290
 QY 210 ALHCGILNNIVYL 222
 DB 291 THH-----NEWMILL 299

RESULT 14
 021884 PRELIMINARY; PRT; 1603 AA.
 AC 021884;
 DT 01-NOV-1996 (TRENBLREL. 01, Created)
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
 DT 01-NOV-1999 (TRENBLREL. 12, Last annotation update)
 DE R09H10.5 PROTEIN.
 GN R09H10.5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderiinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA BASHAM V.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 DT [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94150718.
 RA WILSON R., AINSFORTH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DUBBIN R., FAVELLO A., FULLON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIEREN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., STURSTON J.,
 RA THIERREY-MEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans."
 RL Nature 368:32-38(1994).
 DR EMBL: Z77134; CAB00874.1; -
 DR HSSP: P00750; 1TPG.
 SQ SEQUENCE 1603 AA; 175349 MW; 38FEA309 CRC32;

Query Match 5.1%; Score 82; DB 5; Length 1603;
 Best Local Similarity 21.9%; Pred. No. 5.7;
 Matches 44; Conservative 29; Mismatches 82; Indels 46; Gaps 11;
 QY 9 GMTLPFRSLRRTPTSDRLPFETETATLIEDHPLPCNTLTMHNSVY-----R 63
 DB 972 GFSPDPSPDFRSDLVATYHQPVHIL-----SSLQSDPSIDVSYNNNNNAFTSLFTR 1028
 QY 64 GLPCSVGFTLQEM-----VPMVMTRELYLRKC-----LCCANIDIM---T 110
 DB 1029 TSGCKTEYIMNOETQASAGDTFTMETTTITTEYLTMRORAYCSDILGICINGITLGCN 1088
 QY 111 SMHIGYSWALHCH--CSPGSL-----OC--IAGQVLAFFRMVVDGAMFNORFIW 160
 DB 1089 CQCVNGYTS--LHCEVPTQONGSVVDVDFKCCPSITDG-----DSCQYTTCTWT 1135
 QY 161 YREVVNYNMPKEVMSVFM 181

DB 1136 --NFETHDPREYNFOQIVFV 1154

RESULT 15
097285 PRELIMINARY; PRT; 433 AA.
AC 097285;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE PFC0915W PROTEIN.
GN PFC0915W.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-3D7;
RA OLIVER K., BOWMAN S., HARRIS D., LAWSON D., QUAIL M., BARRELL B.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL034559; CAB39031.1;
DR PROSITE; PS00039; DEAD_ATP_HELICASE; 1.
KW ATP-binding; RNA-binding; Helicase.
SQ SEQUENCE 433 AA; 49413 MW; C419B9EF CRC32;

Query Match 5.18; Score 80.5; DB 5; Length 433;
Best Local Similarity 21.18; Pred. No. 1.9;
Matches 58; Conservative 49; Mismatches 89; Indels 79; Gaps 15;

QY 46 PLPECTLMHNSYVGRGLPCSVGFLIQEWVPMVNLRELVLTRCMHYCLCAN 105
DB 117 PLLEKCT---HNFTQGL-----ILPTRELALQTSAMIKELGKHKVOCMV 161
QY 106 IDMTSM--MIGYESMALHCHCSPSLQICNGQV--LASWFRVVDGA--MENQRF 158
DB 162 ITGISTLRDIDIMRLYN--VHILGTPRIIDLANKDVANLSGCHIVWMDADKLSPF 219
QY 159 IWTREVVNMPKE--VMENSVF--MRGNHLYTLRLMTDGHVGSVPAMSF----GYS 209
DB 220 QPIVEELMKFLPKKQIIMTSATFPVTKERAYLS--DAHEINIMDELTLNGITQY 276
QY 210 A-----LHC-----GIINNTIVLCCSYCADLSEIRVRCARTRRLMLRAVRIIAE 255
DB 277 AFWKROKVAHCLNTLFAKLOINQAIIFCNS-----ITVELLAK 315
QY 256 ETTAMLYSCRTERRRQOFIRA-LIQHHRPIIMHDY 289
DB 316 KITELGYS-----SFYIHARMSQTHRNRYFHD 343

Search completed: June 23, 2000, 10:07:44
Job time: 947 sec

A: Molecule type: DNA
A: Residues: 1-120 <SPR>
A: Cross-references: EMBL:X73487; NID:g313361; PIDD:CAAS1902.1; PTD:g313387
C: Superfamily: adenovirus early E4 13k protein

Query Match	39.9%	Score 248;	DB 2;	Length 120;
Best Local Similarity	40.2%;	Pred. No. 8.4e-20;		
Matches 43;	Conservative 23;	Mismatches 41;	Indels 0;	Gaps 0;

```

0y      1 MVLPAHPAPPVCDSDQNECVGMLGVAISAVVDYIRAAAHGVYIEPEARGHIDALREWITY 60
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1 MPPLCPPPPPSVKDTAAICIAWLGLAHASCVDTLRFIKHDKITPEAEYILASLEWLYE 60

```

61 NYTTERSKRDRRRRSVCHARTWCFRKDYRRSIWHDTTNTISV 107

775449

M.: Shen, M.; Yamachyan, J. J.; Lam, P.; McDonald, L.; Utterback, T.; Zaleski, C.; Smith, H. O.; Venter, J. C.; Fraser, C. M. *Science* 266, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.

A:Reference number: A75250

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-373 <WHI>
A:Cross-references: GB:AE001952; GB:AE000513; NID:96458725; PIDN:AP10581.1; PID:96458737
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1006
A:Map position: 1

Query Match	13.0%	Score 81;	DB 2;	Length 373;
Best Local Similarity	35.3%	Pred. No. 0.31;		
Matches 24;	Conservative 7;	Mismatches 21;	Indels 16;	Gaps 3

```
QY      59 YNRYTER 66
        |::||
Db      305 YRHQVER 312
```

Query Match	12.9%	Score 80;	DB 2;	Length 118;
Best Local Similarity	29.0%	Pred. No. 0.13;		
Matches 29;	Conservative 12;	Mismatches 39;	Indels 20;	Gaps

```

QY 12 CDSQNCVGLWGLVAY-----SAVDVYIRAA-AHEGVYIEPEANGRLDALREWITY 60
      |::|||::|||::|||::|||
Db 11 CANKYKRPGWHYANYGCTCGSGGRTPVDDVDRCCQAHDKCYEDAE---KLGCYPKWITY 67

```

```

QY 61 NYTYS---KRDRRRSVCHAPTW--CFRKIDYVR 94
    ||| : ||| : ||| : ||| : |||
Db 68 NYCGANGPYCKTRTKCFVNCNDVVAADCFASYPNRR 107

```

D64413
cobalamin biosynthesis protein N homolog - *Methanococcus jannaschii*
C:Species: *Methanococcus jannaschii*

A: Authors: Kane, B.P.; Borodovsky, M.; Klek, H.P.; Fraser, C.M.; Smith, H.O.; Wease
A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*
A:Reference number: A64300, M01D:96337999
A:Accession number: D54413

A: molecule type: DNA
A: Residues: 1-1232 <BOLD>
A: Cross-references: GB:U67534; GB:L77117; NID:g1591579; PIRID:AMB98910.1; PID:g1591582
C: Genets:
A: Map position: FOR837021-840719
A: Start codon: TGG
C: Superfamily: Rhodobacter capsulatus magnesium-protoporphyrin O-methyltransferase

Query Match	11.7%	Score 72.5;	DB 2;	length 1232;
Best Local Similarity	23.8%	Pred. No. 8.5;		
Matches 35; Conservative	14;	Mismatches 51;	Indels 47;	Gaps

```

      : | :|: | : |
Db 144 EYEPMPWQIGIYYKGYFETLDD-----Y
      : | :|: | : |
Dv  80 VDVVDDSTLDDTMMY-----MTHNGAV 111

```

S3I511	Ig heavy chain - human
C;Species:	Homo sapiens (man)
Date:	13-Jan-1995 #sequence revision 13-Jan-1005 #cont change 22 Jul 100

Query Match 10.4%; Score 64.5; DB 2; Length 284;
Best Local Similarity 30.4%; Pred. No. 15;
Matches 17; Conservative 9; Mismatches 21; Indels 9; Gaps 2;

OY 63 YTEBSKRRDRRRSCHARTFCFRKYDYVRST-----WHDTTNTISVSASVS 113
DB 21 YVDIDRRKRRSR-----RKSMSRSHGFYERRESTEILKRWTVGVSIVGVAANHV 72

RESULT 11

valyl-tRNA synthetase (vals) PAB125 - Pyrococcus abyssi (strain Orsay)

C/Species: Pyrococcus abyssi

C/Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 24-Nov-1999

C/Accession: G75014

R:anonymous; Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome stru

A:Reference number: A75001

A:Accession: G75014

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-891 <KAM>

A:Cross-references: GB:A248286; GB:AL096836; NID:95458960; PIDN:CAB50557.1; PID:el51645

A:Experimental source: strain Orsay

A:Genetics:

A:Gene: PAB1255

C:Superfamily: valine--tRNA ligase

Query Match 10.3%; Score 64; DB 2; Length 891;
Best Local Similarity 23.1%; Pred. No. 51;
Matches 25; Conservative 16; Mismatches 43; Indels 24; Gaps 5;

OY 4 PALPAPPCDSQNECVG-----WLGVAVSAYVDVIRAAHGGVYIEPEARGLDALREMI 58
DB 435 PVEKCPVCGAIEIPYDVLDCWD-----SITPLITRWHALIGDEA-----KKWF 484

OY 59 YNTYTERSKRRDRRRSVCHARTW---FCFRKYDYVRSTIWDHTTN 103
DB 485 EHNFPPTA-----LRFOTDIIITWAFYITRFTKLGKKRWKNIIVIN 526

RESULT 12

T00790 10.3%; Score 64; DB 2; Length 891;
Best Local Similarity 23.1%; Pred. No. 51;
Matches 25; Conservative 16; Mismatches 43; Indels 24; Gaps 5;

ubiquitin-specific proteinase homolog F24L7.8 - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cross)

C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 30-Apr-1999

C/Accession: T00790

R:Ronsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul

submitted to the EMBL Data Library, February 1998

A:Description: Arabidopsis thaliana chromosome II BAC F24L7 genomic sequence.

A:Reference number: Z14204

A:Accession: T00790

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1083 <KOD>

A:Cross-references: EMBL:AC003974; NID:92914688; PID:92914695

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map-position: 2

A:Introns: 416/2; 1004/2

Query Match 10.3%; Score 64; DB 2; Length 1083;
Best Local Similarity 24.3%; Pred. No. 62;
Matches 18; Conservative 12; Mismatches 30; Indels 14; Gaps 1;

OY 14 SQNECVGLVAVSAVVDVIRAAHGGVYI-----EPEARGLDALREMI 59
DB 931 SDNECEWMDLAVSEVIVYKRDARKKVLINKAPVLTIIHKRFSGDARGSVKSLSGHD 990

OY 60 YNTYTERSKRRDR 73
DB 991 FQETIDSKYMDTR 1004

RESULT 13

A56921 kinesin family protein KIF1A - mouse

C/Species: Mus musculus (house mouse)

C/Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999

C/Accession: A56921

R:Okada, Y.; Yamazaki, H.; Sekine-Aizawa, Y.; Hirokawa, N.

Cell 81, 769-780, 1995

A:Title: The neuron-specific kinesin superfamily protein KIF1A is a unique monomeric

A:Reference number: A56921; M0ID:95292344

A:Accession: A56921

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1695 <RES>

A:Cross-references: GB:D29951; NID:9976234; PIDN:BA06221.1; PID:9976235

C:Superfamily: kinesin-related protein unc-104; kinesin motor domain homology; plects

C:Keywords: P-loop

F/6-360/Domain: kinesin motor domain homology <KNOT>

F/97-104/Region: nucleotide-binding motif A (P-loop)

Query Match 10.3%; Score 64; DB 2; Length 1695;
Best Local Similarity 23.4%; Pred. No. 97;
Matches 32; Conservative 19; Mismatches 56; Indels 30; Gaps 7;

OY 7 PAPPV--CDSQNECVGLV-----AYSADVIRAAHGGVYIEPEAR---GRLD 52
DB 613 PAEPVMAAQLLEKQIDKQMEGRLOLEQYRREPEARLYLEQRLDESKLE 672

OY 53 ALREMIYTYTERSKRRDRRRSV-----CHARTWCFRKYDY-----VRSIMDT 101
DB 673 ALQKQMSRYEYVEVEEPEDEYQWTERECALM-AFRKMKYQVTSJRDLLMGVAI 731

OY 101 ---TNTISVSASVS 114
DB 732 FLKENAIVSEVKKKVQ 748

RESULT 14

T16188 10.2%; Score 63.5; DB 2; Length 264;
Best Local Similarity 23.2%; Pred. No. 17;
Matches 26; Conservative 17; Mismatches 48; Indels 21; Gaps 6;

hypothetical protein F27D9.6 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C/Accession: T16188

R:Bentley, D.

submitted to the EMBL Data Library, February 1996

A:Description: The sequence of C. elegans cosmid F27D9.

A:Reference number: Z18473

A:Accession: T16188

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-264 <BEN>

A:Cross-references: EMBL:U49829; NID:91203924; PID:91203928; PIDN:AAA93385.1; CESP:F2

A:Gene: CESP:F27D9.6

A:Introns: 28/2; 57/3; 117/2; 173/3; 231/2

Query Match 10.2%; Score 63.5; DB 2; Length 264;
Best Local Similarity 23.2%; Pred. No. 17;
Matches 26; Conservative 17; Mismatches 48; Indels 21; Gaps 6;

OY 13 DSQNECV---GW---LGVAVSAYVDVIRAAHGGVYIEPEARGLDALREMIYNYT 64
DB 6 DGHIVCVSINGSGEMGSLCYCTSKFAVNGAM-ESLQMLRDRG-LEGIKITTLTYFYFA 63
OY 65 ESKRRDRRRSVCHARTWCF-----KKYDYVRSTIWDHTTNTISV 108

Db 64 RFPMLNNMPTC--TWPFMSIRSCSNMVDILKEKHAVPSTITLI 112

RESULT 15

T40584 Probable involvement in ergosterol biosynthesis - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T40584

R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, January 1999

A:Reference number: 221938

A:Accession: T40584

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-516 <SEE>

A:Cross-references: EMBL:AL035216; PIDN:CNA22812.1; GSPDB:GN00067; SPDB:SPBC646.08c

A:Experimental source: strain 972h-; cosmid c646

C:Genetics:

A:Gene: SPDB:SPBC646.08c

A:Map position: 2

Query Match

Best Local Similarity 22.9% Pred. No. 34;

Matches 25; Conservative 14; Mismatches 47; Indels 23; Gaps 4;

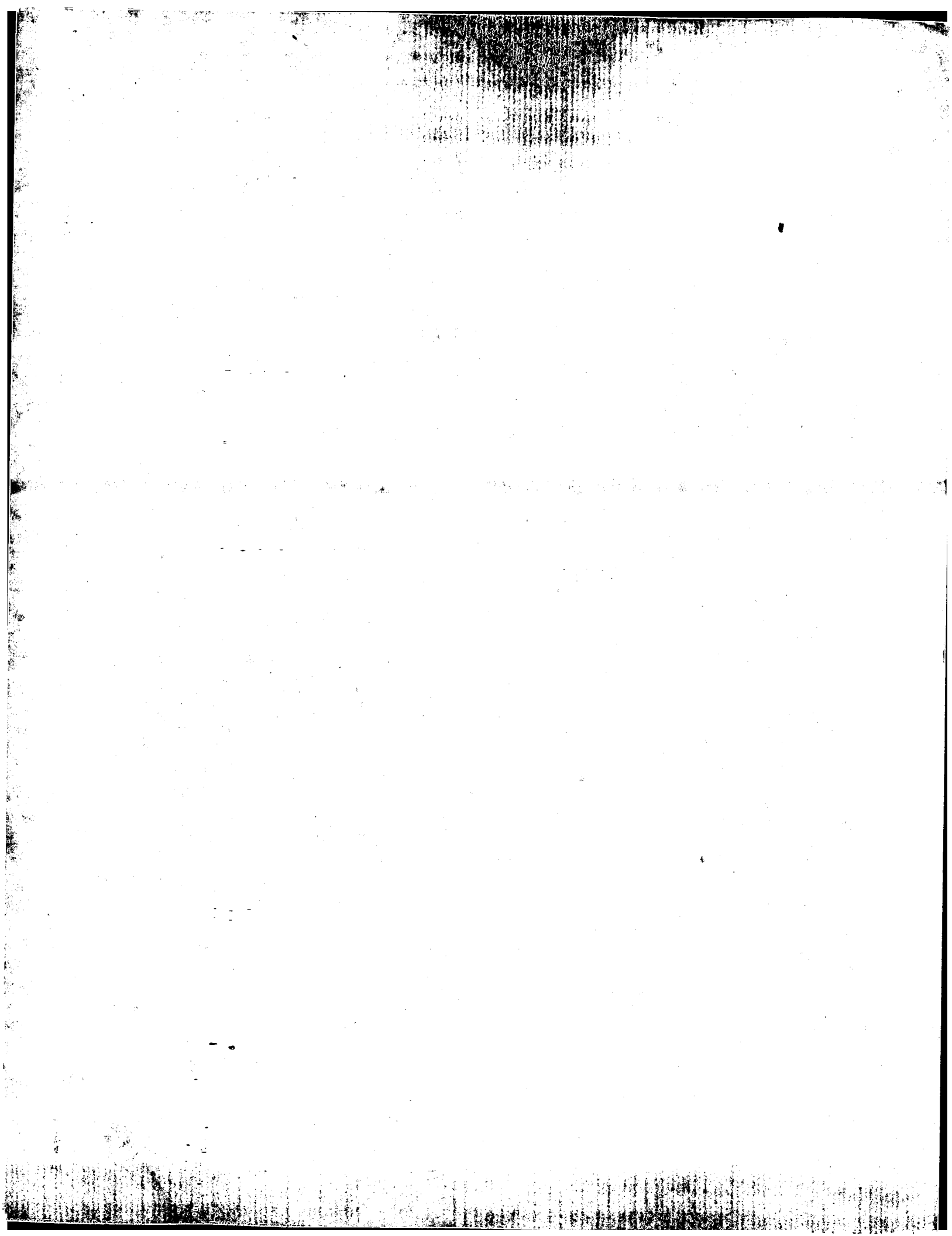
21 WLGVASAV-----VDVRAAAHSGVTEPEARGRLDALREMYIYNYTTERS 67

Db 354 WLGKPRSLVEGVGYRYPDSNDTIDSIKAVPKENILATFKGN-----WRNCIFYSAGESE 408

68 KRR--DRRRSVCHARTWCFRKDYVRSIMHDTTNTISVSAHSVQ 114

Db 409 SRLVDLNLDELVHKRCPPIDKQFPESRKIMPPVTHN---ILAKHYIQ 454

Search completed: June 23, 2000, 10:05:47
Job time: 1752 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2000, 10:14:48 ; Search time 13.99 Seconds
(without alignments)

248.168 Million cell updates/sec

Title: US-09-214-478-4

Perfect score: 621
Sequence: 1 MYLPALPAPVCDSONECVGV.....\$IWHDTTNTISVSAHSVQ 114

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 83857 seqs, 30454973 residues

Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_38.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	618	99.5	E413_ADE02	P03240 human adeno
2	248	39.9	E413_ADE12	P36709 human adeno
3	80	12.9	PA2A_PSETE	P23026 pseudonata
4	68	11.0	RP54_RHIME	P17263 rhizobium m
5	67	10.8	YO11_BPL2	P42346 bacterioph
6	66	10.6	RRPL_AHSV9	O70695 african hor
7	65.5	10.5	PAPA_CAVPO	P70683 c platelet-
8	64	10.3	KFLA_HUMAN	Q12756 homo sapien
9	64	10.3	KFLA_MOUSE	P33173 mus musculu
10	63	10.1	MOG_MOUSE	O61885 mus musculu
11	63	10.1	DAPF_STRCO	O69869 streptomyc
12	63	10.1	CEGR_HUMAN	Q16739 homo sapien
13	63	10.1	PSD3_DAVCA	O06364 danus caro
14	63	10.1	DYHC_YEAST	P36022 saccharomyc
15	62	10.0	MOG_RAT	O63345 rattus norv
16	62	10.0	EXG_YAPLI	Q12725 yarrowia 11
17	62	10.0	STAR_DROME	P42519 drosophila
18	61.5	9.9	UL87_HCMVA	P16730 human cytom
19	61	9.8	VP9_MTV	P12326 wound tumor
20	61	9.8	PARF_SALTY	P31398 salmonella
21	61	9.8	UL52_HCMVA	P16793 human cytom
22	61	9.8	RPO2_CAVPK	P16716 caeaploxytr
23	60.5	9.7	Y4YR_RHISN	P55726 rhizobium s
24	60	9.7	UL66_HCMVA	P16822 human cytom
25	60	9.7	VIF_HV2ST	P20878 human immun
26	60	9.7	ARGA_PSEAE	P22567 pseudomonas
27	60	9.7	COBL_PSEAE	P21921 pseudomonas
28	59.5	9.6	LCB3_ROBPS	O41160 scilina pse
29	59.5	9.6	POLL_SCRIC	O03277 scilina copr
30	59.5	9.6	STL_TREPA	O83466 treponema p
31	59.5	9.6	YLJ2_CAEEL	P34367 caenorhabdl
32	59	9.5	HLJC_HUMAN	P01744 homo sapien
33	59	9.5	MOG_BOVIN	P55803 bos taurus
34	59	9.5	VP9_MTVNJ	P31611 wound tumor

35	59	9.5	339	1	YGOB_HAEIN	P44039 haemophilus
36	58.5	9.4	445	1	SPRI_YEAST	P32603 saccharomyc
37	58.5	9.4	833	1	CM41_YEAST	P53008 saccharomyc
38	58	9.3	185	1	VCOT_ADE40	O89532 human adeno
39	58	9.3	215	1	VIF_HV2G1	P18043 human immun
40	58	9.3	271	1	YK23_YEAST	P36136 saccharomyc
41	58	9.3	358	1	YK15_CAEEL	P34316 caenorhabdl
42	58	9.3	426	1	YJIN_ECOLI	P39385 escherichia
43	58	9.3	508	1	COB1_MYCTU	Q10677 mycobacteri
44	57.5	9.3	182	1	LEC7_CAEEL	O09605 caenorhabdl
45	57.5	9.3	251	1	HXB4_FUGRU	O13074 fugu rubrip

ALIGNMENTS

RESULT 1	E413_ADE02	STANDARD:	PRT:	114 AA.
ID	E413_ADE02	STANDARD:	PRT:	114 AA.
AC	P03240;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	01-APR-1990 (Rel. 14, Last annotation update)			
DE	PROBABLE EARLY E4 13 KD PROTEIN.			
OS	Human adenovirus type 2.			
CC	Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.			
CC	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 82059444.			
RA	Hershe J., Rigdout M., Dupont de Dinechin S., Galibert F.,			
RT	"Nucleotide sequence of adenovirus 2 DNA fragment encoding for the			
RI	carboxylate region of the fiber protein and the entire E4 region."			
RL	Nucleic Acids Res. 9:4023-4042(1981).			
CC	-1- MISCELLANEOUS: THIS PROBABLE PROTEIN WAS ASSIGNED BY CORRELATING			
CC	EM DATA AND S1 DIGESTION STUDIES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL, J01917; -, NOT_ANNOTATED_CDS.			
DR	PIR; A03806; Q4AD2.			
KW	Early protein.			
SO	SEQUENCE 114 AA; 13321 MW; 72FB72749D563457 CRC64;			
Query Match	99.5%; Score 618; DB 1; Length 114;			
Best Local Similarity	99.1%; Pred. No. 1e-61;			
Matches 113; Conservative 1; Mismatches 0; Indels 0; Gaps 0;				
OY	1 MYLPALPAPVCDSONECVGVSAVAVDIYRAAHEGYIEFEANGRLDALREWTYY 60			
DB	1 MYLPALPAPVCDSONECVGVSAVAVDIYRAAHEGYIEFEANGRLDALREWTYY 60			
OY	61 NYTTERSKRRDRRRRSVCHARTWFCERKDYVRSIMHDTTNTISVSAHSVQ 114			
DB	61 NYTTERAKRRDRRRRSVCHARTWFCERKDYVRSIMHDTTNTISVSAHSVQ 114			
RESULT 2	E413_ADE12	STANDARD:	PRT:	120 AA.
ID	E413_ADE12	STANDARD:	PRT:	120 AA.
AC	P36709;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	01-JUN-1994 (Rel. 29, Last annotation update)			
DE	PROBABLE EARLY E4 13 KD PROTEIN.			
OS	Human adenovirus type 12.			
CC	Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.			

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94076430.
 RA Sprengel J., Schmitz B., Heuss-Neltzel D., Zock C., Doerfler W.;
 RT "Nucleotide sequence of human adenovirus type 12 DNA: comparative
 RT functional analysis.";
 RL J. Virol. 68:379-389(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 90272430.
 RA Hogenkamp T., Esche H.;
 RT "Nucleotide sequence of the right 10% of adenovirus type 12 DNA
 RT encoding the entire region E4.";
 RL Nucleic Acids Res. 18:3065-3066(1990).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X73487; CAAS1902.1;
 DR EMBL: X51800; CAB57853.1;
 DR PIR: S33953; S33953.
 DR PIR: S10865; S10865.
 KW Early Protein.
 SQ SEQUENCE 120 AA; 13600 MW; 2004EAA61B792AF CMC64;
 Query Match 39.9%; Score 248; DB 1; Length 120;
 Best Local Similarity 40.2%; Pred. No. 1e-20;
 Matches 43; Conservative 23; Mismatches 41; Indels 0; Gaps 0;
 QY 1 MYPLPAPPCVDSQNECYGMVGVAVYVYDRAAAHGVYIEPARGDALREWTYY 60
 DB 1 MFLPCLPPPPVSRDTRACIAMGLAHASCVDLRFKHKHDKITPEARYTILASIREWLYF 60
 QY 61 NYTTERSKRRDRRRRSVCHARTFCFRKYDYRRRSIMHDTTNTISV 107
 DB 61 ALTERORCKOKGKRGALTSGRTWFCFFYEDARKSVYDARQVYSL 107
 RESULT 3
 PAZA_PSETE STANDARD; PRT; 118 AA.
 AC P23026;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PHOSPHOLIPASE A2 HOMOLOG, TEXTILOTOXIN A CHAIN (EC 3.1.1.4).
 OS Pseudonaja textilis (Eastern brown snake).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Lepidosaurota;
 OC Squamata; Scleroglossa; Serpentes; Colubroidae; Elapidae;
 OC Acanthophiline; Pseudonaja.
 RN [1]
 RP SEQUENCE.
 RX TISSUE-VENOM.
 RX MEDLINE: 93160241.
 RA Pearson J.A., Tyler M.I., Retson K.V., Howden M.E.H.;
 RT "Studies on the subunit structure of textilotoxin, a potent
 RT presynaptic neurotoxin from the venom of the Australian common brown
 RT snake (Pseudonaja textilis). 3. The complete amino-acid sequences of
 RT all the subunits.";
 RL Blochim. Biophys. Acta 1161:223-229(1993).
 RN [2]
 RP SEQUENCE.
 RX TISSUE-VENOM.
 RX MEDLINE: 88000715.
 RA Tyler M.I., Barnett D., Nicholson P., Spence I., Howden M.E.H.;
 RT "Studies on the subunit structure of textilotoxin, a potent
 RT neurotoxin from the venom of the Australian common brown snake

RT (Pseudonaja textilis).";
 RL Blochim. Biophys. Acta 915:210-216(1987).
 CC -1- FUNCTION: POTENT PRESYNAPTIC NEUROTOXIN POSSESSING PHOSPHOLIPASE
 CC ACTIVITY. SUBUNIT A IS LETHAL TO MICE AT 4 MG/KG (I.V.). IT IS
 CC ESSENTIAL FOR THE NEUROTOXICITY OF TEXTILOTOXIN. SUBUNIT A
 CC POSSESSES A LOW PHOSPHOLIPASE ACTIVITY.
 CC -1- FUNCTION: PA2 CATALYZES THE CALCIUM-DEPENDENT HYDROLYSIS OF THE
 CC 2-ACYL GROUPS IN 3-SN-PHOSPHOGLYCERIDES.
 CC -1- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + H(2)O = 1-ACYLGLYCERYL-
 CC PHOSPHOCHOLINE + A FATTY ACID ANION.
 CC -1- SUBUNIT: PENTAMER OF THREE NON-COVALENTLY LINKED SUBUNITS A, B,
 CC AND C, AND TWO IDENTICAL COVALENTLY LINKED D SUBUNITS.
 CC -1- MISCELLANEOUS: ALL SUBUNITS ARE NECESSARY FOR MAXIMUM LETHALITY.
 CC LD(50) IS 0.001 MG/KG BY INTRAPERITONEAL INJECTION.
 CC -1- SIMILARITY: BELONGS TO THE PHOSPHOLIPASE A2 FAMILY.
 DR PIR: S29651; S29651.
 DR HSP: P00609; 2NOT.
 DR PRINTS: PR00389; PHPLIPASEA2.
 DR PROSITE: PS00118; PA2_HIS; 1.
 DR PROSITE: PS00119; PA2_ASP; 1.
 DR PRAM: PF000068; Phoslip; 1.
 KW Hydrolase; Lipid degradation; Calcium; Multigene family; Venom;
 FT ACT_SITE 48 48 BY SIMILARITY.
 FT ACT_SITE 92 92 BY SIMILARITY.
 FT DISULFID 11 71 BY SIMILARITY.
 FT DISULFID 27 117 BY SIMILARITY.
 FT DISULFID 29 45 BY SIMILARITY.
 FT DISULFID 44 98 BY SIMILARITY.
 FT DISULFID 51 91 BY SIMILARITY.
 FT DISULFID 60 84 BY SIMILARITY.
 FT DISULFID 78 89 BY SIMILARITY.
 SQ SEQUENCE 118 AA; 13849 MW; 3B43FBC0A751C09E CMC64;
 Query Match 12.9%; Score 80; DB 1; Length 118;
 Best Local Similarity 29.0%; Pred. No. 0.041;
 Matches 29; Conservative 12; Mismatches 39; Indels 20; Gaps 5;
 QY 12 CDSQNCVGMVGVAVY-----SAVDYVRAA-AHEGYIEPARGDALREWTYY 60
 DB 11 CANKRYRPPGMHANYANGCYGSGRGTPVDVDRCCQADHKCYEDAE--KLGCPKPTTY 67
 QY 61 NYTTERSKRRDRRRRSVCHARTFCFRKYDYRRRSIMHDTTNTISV 94
 DB 68 NYTCGANGPYCKTRTKCQRFVNCVVAADCASTPIYNR 107
 RESULT 4
 RP54_RHIME STANDARD; PRT; 514 AA.
 ID RP54_RHIME
 AC P17263; 059751;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 15-FEB-2000 (Rel. 39, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE RNA POLYMERASE SIGMA-54 FACTOR (VERSION 1).
 GN RPON OR NTRA.
 OS Rhizobium meliloti.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 87222158.
 RA Ronson C.W., Nixon B.T., Albright L.M., Ausubel F.M.;
 RT "Rhizobium meliloti ntra (rpon) gene is required for diverse
 RT metabolic functions.";
 RL J. Bacteriol. 169:2424-2431(1987).
 RN [2]
 RP REVISIONS TO C-TERMINUS.
 RX MEDLINE: 89197760.
 RA Albright L.M., Ronson C.W., Nixon B.T., Ausubel F.M.;
 RT "Identification of a gene linked to Rhizobium meliloti ntra whose
 RT product is homologous to a family to ATP-binding proteins.";

RL J. Bacteriol. 171:1932-1941(1989).

CC -1- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
CC THEN IS RELEASED.

CC -1- FUNCTION: THIS SIGMA FACTOR IS RESPONSIBLE FOR THE EXPRESSION OF
CC THE NITROGEN FIXATION PROTEINS (NIF OPERON), GUN4 AND DCA FOR
CC DICARBOXYLATE TRANSPORT. THE OPEN COMPLEX (SIGMA-54 AND CORE RNA
CC POLYMERASE) SERVES AS THE RECEPTOR FOR RECEIPT OF THE MELTING
CC SIGNAL FROM THE REMOTELY BOUND ACTIVATOR PROTEINS NIFA, NTRC, OR
CC DCD FOR THE EXPRESSION OF THE REGULATED PROTEINS.

CC -1- SIMILARITY: BELONGS TO THE SIGMA-54 FACTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M16513; AAD15117.1; -
DR EMBL: M24926; AAA26349.1; -
DR PRINTS; PR00045; SIGMA54.FCT.
DR PROSITE; PS00717; SIGMA54_1; 1.
DR PROSITE; PS00718; SIGMA54_2; 1.
DR PROSITE; PS50044; SIGMA54_3; 1.
DR TRASN:~PF00309; Sigma54_factors; 1.
KW Transcription regulation; Sigma factor; DNA-directed RNA polymerase;
KW DNA-binding; Nitrogen fixation.
FT DOMAIN 11 29 GLN-RICH.
FT DOMAIN 21 42 LEUCINE-ZIPPER (POTENTIAL).
FT DOMAIN 189 210 LEUCINE-ZIPPER (POTENTIAL).
FT DNA_BIND 393 412 H-T-H MOTIF (POTENTIAL).
FT SITE 482 490 RPOB BOX.
FT CONFLICT 509 514 AKAGF -> PRPDRSCQQAASA (IN REF. 1).
SQ SEQUENCE 514 AA; 36678 MW; C291DAE1DD2D034C CRC64;

Query Match 11.0%; Score 68; DB 1; Length 514;
Best Local Similarity 28.4%; Pred. No. 4.3;
Matches 27; Conservative 20; Mismatches 36; Indels 12; Gaps 4;

QY 23 GVAVSAVDVIRAAHEG---VYIEPEARGLDALREMIYNYTNSKRDRRRRSVCH 79
ID 291 GVFAIIPDVVRAAPDGMVLEINPDLPRV-----LVNHDYETLS--RSSKNSGEQ 343
AC P42546; STANDARD; PRT; 218 AA.

QY 80 ARTFECFRKYDVRRSTHDTTWTISVSHSVQ 114
DB 344 AFINECLONAMWTRSL--DQRANTMKVASEIVR 376

RESULT 5
Y011_BPL2 STANDARD; PRT; 218 AA.
AC P42546;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHEICAL 25.6 KD PROTEIN (ORF11).
OS Bacteriophage L2.
OC Viruses; dsDNA viruses, no RNA stage; Plasmaviridae; Plasmavirus.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE; 94215869.
RA Maniloff J., Kampo G.J., Dascher C.C.;
RT "Sequence analysis of a unique temperature phage: mycoplasma virus
RL L2.";
RL Gene 141:1-8(1994).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: I13696; AAA87967.1; -
KW Hypothetical protein.
SQ SEQUENCE 218 AA; 25564 MW; A6DD8C582B8EFCFE CRC64;

Query Match 10.8%; Score 67; DB 1; Length 218;
Best Local Similarity 24.3%; Pred. No. 2.2;
Matches 17; Conservative 12; Mismatches 31; Indels 10; Gaps 2;

QY 14 SONECVGMLGVASVAVDVIRAAHEG---VYIEPEARGLDALREMIYNYT 63
DB 47 SDELIGKMGCSRSGLYDAFVLERGLIKRIYSDTKRKNRTEIERNIDLAASWKLITLY 106
QY 64 TERSKRDRR 73
DB 107 DKAYKNAPRR 116

RESULT 6
RRPL_AHSV9 STANDARD; PRT; 1305 AA.
ID RRPL_AHSV9
AC 070695;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (YPL).
GN SI.
OS African horse sickness virus 9 (AHSV-9) (African horse sickness virus
OS (serotype 9)).
OS Viruses; dsRNA viruses; Reoviridae; Orbivirus.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE; 98202706.
RA Vreede F.T., Huismans H.;
RT "Sequence analysis of the RNA polymerase gene of African horse
RL Arch. Virol. 143:413-419(1998).
CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE -> N PYROPHOSPHATE
CC + RNA(N).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL: U94887; AAC40586.1; -
KW Transferase; Nucleotidyltransferase; RNA-directed RNA polymerase.
SQ SEQUENCE 1305 AA; 150293 MW; 321B9E7F3CF6F11A CRC64;

Query Match 10.6%; Score 66; DB 1; Length 1805;
Best Local Similarity 30.2%; Pred. No. 20;
Matches 19; Conservative 9; Mismatches 21; Indels 14; Gaps 3;

QY 49 GRDLALREMIYNYTNSKRDRRRRSVCHARR--WFCRRKVD-----YVRSI 96
DB 22 GRIDGTR-MYEEYRYSKWRRTRRKGTGYTDEFLERDAGRLKLYDLOVIREAS 79
QY 97 WHD 99
DB 80 WED 82

RESULT 7
PAPA_CAVPO STANDARD; PRT; 436 AA.
ID PAPA_CAVPO

AC P70683; 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE PLATELET-ACTIVATING FACTOR ACETYLHYDROLASE PRECURSOR (EC 3.1.1.47)
 DE (PAF ACETYLHYDROLASE) (PAF 2-ACTIHYDROLASE) (LID-ASSOCIATED
 DE PHOSPHOLIPASE A2) (LID-PLA(2)) (2-ACETYL-1-ALKYLGLYCEROPHOSPHOCHOLINE
 DE ESTERASE) (1-ALKYL-2-ACETYLGLYCEROPHOSPHOCHOLINE ESTERASE).
 GN PLA2G7.
 OS Cavia porcellus (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Hystriocognathi; Caviidae; Cavia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HARTLEY; TISSUE-LIVER;
 RC MEDLINE; 97103479.
 RA Karsawa K., Kuge O., Kawasaki K., Nishijima M., Nakano Y., Tomita M.,
 RA Yokoyama K., Setaka M., Nojima S.;
 RT "Cloning, expression and characterization of plasma
 RT platelet-activating factor acetylhydrolase from guinea pig.";
 RL J. Biochem. 120:838-844(1996).
 CC -1- FUNCTION: MODULATES THE ACTION OF PLATELET-ACTIVATING FACTOR (PAF)
 CC BY HYDROLYZING THE SN-2 ESTER BOND TO YIELD THE BIOLOGICALLY
 CC INACTIVE LYSO-PAF. HAS A SPECIFICITY FOR SUBSTRATES WITH A SHORT
 CC RESIDUE AT THE SN-2 POSITION. IT IS INACTIVE AGAINST LONG-
 CC CHAIN PHOSPHOLIPIDS.
 CC -1- CATALYTIC ACTIVITY: 2-ACETYL-1-ALKYL-SN-GLYCERO-3-PHOSPHOCHOLINE +
 CC H(2O) -> 1-ALKYL-SN-GLYCERO-3-PHOSPHOCHOLINE + ACETATE.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- TISSUE SPECIFICITY: PLASMA.
 CC -1- SIMILARITY: PARTIAL WITH OTHER LIPASES (PANCREATIC, GASTRIC,
 CC HEPATIC, LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D67037; BAA11054.1;
 DR PROSITE; PS00120; LIPASE_SER; 1.
 KW Hydrolase; lipid degradation; glycoprotein; signal.
 FT SIGNAL 1 21
 FT CHAIN 22 436
 FT ACT_SITE 271 271
 FT ACT_SITE 294 294
 FT ACT_SITE 349 349
 FT CARBOHD 76 76
 FT CARBOHD 200 200
 FT CARBOHD 324 324
 FT SEQUENCE 436 AA; 49062 MW; C359D96E392FEF11 CRC64;

Query Match 10.5%; Score 65.5; DB 1; Length 436;
 Best Local Similarity 28.2%; Pred. No. 6.8;
 Matches 20; Conservative 10; Mismatches 32; Indels 9; Gaps 3;

QY 22 LGVATSAVDVIRAAHE-----GYIIEPEARGLDALRENTY---NYIERSKRRDR 73
 DB 163 IGIELASHGFIYAANVHEHDESAATYTFQDAPAAISGRSMYIYGVGLTEBERKRLAQ 222
 QY 74 RRSVC-HARTW 83
 DB 223 RGECSQALSW 233

RESULT 8
 ID KFLA_HUMAN STANDARD; PRT; 1690 AA.
 AC 012756;

DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE KINISIN-LIKE PROTEIN KIFIA (AXONAL TRANSPORTER OF SYNAPTIC VESICLES).
 GN KIFIA OR ATSV.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RC MEDLINE; 96299637.
 RA Furlong R.A., Zhou C.Y., Ferguson-Smith M.A., Affara N.A.;
 RT "Characterization of a kinesin-related gene ATSV, within the tuberous
 RT sclerosis locus (TSC1) candidate region on chromosome 9q34.";
 RL Genomics 33:421-429(1996).
 CC -1- FUNCTION: MOTOR FOR ANTEROGRADE AXONAL TRANSPORT OF SYNAPTIC
 CC VESICLE PRECURSORS (BY SIMILARITY).
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE KINISIN-LIKE PROTEIN FAMILY. UNC-104
 CC SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 1 FHA DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X90840; CAA62346.1;
 DR HSP; P17119; 3KAR.
 DR MIM; 601255;
 DR PRINTS; PR00380; KINISINHEAVY.
 DR PROSITE; PS00411; KINISIN_MOTOR_DOMAIN1; 1.
 DR PROSITE; PS50067; KINISIN_MOTOR_DOMAIN2; 1.
 DR PROSITE; PS50003; PH_DOMAIN; 1.
 DR PROSITE; PS50006; FHA_DOMAIN; 1.
 DR PFAM; PF00169; PH; 1.
 DR PFAM; PF00225; kinesin; 1.
 DR PFAM; PF00498; FHA; 1.
 KW Motor protein; Microtubules; ATP-binding; coiled coil.
 FT DOMAIN 1 361
 FT DOMAIN 366 383
 FT DOMAIN 429 462
 FT DOMAIN 516 572
 FT DOMAIN 622 681
 FT DOMAIN 801 822
 FT DOMAIN 1575 1673
 FT NP_BIND 97 104
 FT SEQUENCE 1690 AA; 191083 MW; D8DDDEC784624FBAD CRC64;

Query Match 10.3%; Score 64; DB 1; Length 1690;
 Best Local Similarity 23.4%; Pred. No. 43;
 Matches 32; Conservative 19; Mismatches 56; Indels 30; Gaps 7;

QY 7 PAPVY-CDSONECVGLGV-----ASAVVDVIRAAHGVYIEPAR-----GRD 52
 DB 613 PAEVDYDAFQRELEKQGDIMKQEMORLQLEDDYRRERERATYLLQQRIDYRSKLE 672
 QY 53 ALREMIYNYIYTERSKRRDRRRRSV-----CHARTWCFRRKYD-----VRSIWHDT 101
 DB 673 ALQKQMSRYIYVNEDEEPEDEVOQTERECCLALM-AFRKKWKQFISRLDINGNAI 731
 QY 101 ----TNTIYVSAHSVQ 114
 DB 732 ELKEANAISVELKKVQ 748

RESULT 9

KIFIA_MOUSE STANDARD; PRT; 1695 AA.

AC: P33173; 061770; 27, Created

DT: 01-OCT-1993 (Rel. 35, Last sequence update)

DT: 01-NOV-1997 (Rel. 35, Last sequence update)

DT: 15-JUL-1998 (Rel. 36, Last annotation update)

DE: KINESIN-LIKE PROTEIN KIFIA.

GN: KIFIA OR KIFL.

OS: Mus musculus (Mouse).

OC: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

RN: [1]

RP: SEQUENCE FROM N.A.

RC: TISSUE-BRAIN.

RX: MEDLINE; 95292344.

RA: Okada Y., Yamazaki H., Sekine-Alzawa Y., Hirokawa N.;

RT: "The neuron-specific kinesin superfamily protein KIFIA is a unique monomeric motor for anterograde axonal transport of synaptic vesicle precursors."

RT: Cell 81:769-780(1995).

RL: [2]

RP: PRELIMINARY SEQUENCE OF 100-247 FROM N.A.

RC: TISSUE-BRAIN.

RX: MEDLINE; 93077686.

RA: Alzawa H., Sekine Y., Takemura R., Zhang Z., Nangaku M., Hirokawa N.;

RT: "Kinesin family in murine central nervous system."

RT: J. Cell Biol. 119:1287-1296(1992).

RL: [1]

CC: -1- FUNCTION: MOTOR FOR ANTEROGRADE AXONAL TRANSPORT OF SYNAPTIC VESICLE PRECURSORS.

CC: -1- SUBUNIT: MONOMER.

CC: -1- TISSUE SPECIFICITY: EXPRESSED ALMOST EXCLUSIVELY IN ADULT BRAIN TISSUE (MAINLY IN THE CEREBELLUM AND CEREBRUM) WITHIN A SINGLE TYPE OF NEURONAL CELL. WITHIN THE NEURONAL CELL LEVELS ARE CONCENTRATED AROUND THE AXON, WITH SMALLER AMOUNTS IN THE PERINUCLEAR AND SYNAPTIC REGIONS.

CC: -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. UNC-104 SUBFAMILY.

CC: -1- SIMILARITY: CONTAINS 1 FHA DOMAIN.

CC: -1- SIMILARITY: CONTAINS 1 PH DOMAIN.

CC: -----

CC: This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation, the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/or_send_an_email_to_license@sib-sib.ch).

CC: -----

DR: EMBL; D29951; BA06221.1; -

DR: PIR; E44259; E44259.

DR: HSSP; P17119; 3KAR.

DR: MGD; MG1:108391; KIFIA.

DR: PRINTS; PR00380; KINESINHEAVY.

DR: PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.

DR: PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.

DR: PROSITE; PS50003; PH_DOMAIN1; 1.

DR: PROSITE; PS50006; FHA_DOMAIN1; 1.

DR: PFM; PFM0169; PH; 1.

DR: PFM; PFM00225; kinesin; 1.

DR: PFM; PFM00498; FHA; 1.

KW: Motor protein; Microtubules; ATP-binding; coiled coil.

FT: DOMAIN 1 361

FT: DOMAIN 366 383

FT: DOMAIN 429 462

FT: DOMAIN 516 572

FT: DOMAIN 622 681

FT: DOMAIN 801 822

FT: DOMAIN 1580 1678

FT: NP_BIND 97 104

SO: SEQUENCE 1695 AA; 191724 MW; DEC3B8BC9C66 CRC64;

Query Match 10.3%; Score 64; DB 1; Length 1695;

Best Local Similarity 23.4%; Pred. No. 44;

Matches: 32; Conservative 19; Mismatches 56; Indels 30; Gaps 7;

QY 7 PAPPV-CDSONECVGMWV-----AYSADVIRAAHEGVIEEAR---GRID 52

DB 613 PAEPVMAFQRELKQDGNKQMEQKLOELEDYRREAREATLLQQRUDYSKLE 672

QY 53 ALREWIYNYTERSKRRDRRSV-----CHARTWCFCKRYD-----VRSIWHDT- 101

DB 673 ALKQMSDRYPPEVNEEPEDEVOETRECELAIW-AFRKKWYQFTSLRDLWGMAI 731

QY 101 ---TNTISVSAHSVQ 114

DB 732 FLKANAISVELKKRYQ 748

RESULT 10

MOG_MOUSE STANDARD; PRT; 246 AA.

ID: MOG_MOUSE P70364;

AC: 061885; Q62003; P70364;

DT: 01-NOV-1997 (Rel. 35, Created)

DT: 01-NOV-1997 (Rel. 35, Last sequence update)

DT: 01-NOV-1997 (Rel. 35, Last annotation update)

DE: MYELIN-OLIGODENDROCYTE GLYCOPROTEIN PRECURSOR.

GN: MOG.

OS: Mus musculus (Mouse).

OC: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

RN: [1]

RP: SEQUENCE FROM N.A.

RC: STRAIN-129;

RX: MEDLINE; 95130110.

RA: Daubas P., Pham-Dinh D., Dautigny A.;

RT: "Structure and polymorphism of the mouse myelin/oligodendrocyte glycoprotein gene."

RT: Genomics 23:36-41(1994).

RL: [2]

RP: SEQUENCE FROM N.A.

RA: Gardiner M.V., Matthieu J.M.;

RT: "Murine and human MOG are highly conserved: cDNA analysis."

RT: Trans. Am. Soc. Neurochem. 24:234-234(1993).

RL: [3]

RP: SEQUENCE OF 29-246 FROM N.A.

RC: STRAIN-BALB/C; TISSUE-BRAIN;

RX: MEDLINE; 93376728.

RA: Pham-Dinh D., Matel M.-G., Nussbaum J.-L., Rousset G., Pontarotti P., Roedel N., Mather I.H., Artzt K., Landahl K.F., Dautigny A.;

RT: "Myelin/oligodendrocyte glycoprotein is a member of a subset of the immunoglobulin superfamily encoded within the major histocompatibility complex."

RT: Proc. Natl. Acad. Sci. U.S.A. 90:7990-7994(1993).

RL: [4]

RP: SEQUENCE OF 29-54.

RC: STRAIN-BALB/C; TISSUE-BRAIN;

RX: MEDLINE; 92218912.

RA: Amalguet P., Gardiner M.V., Zanetta J.-P., Matthieu J.-M.;

RT: "Purification and partial structural and functional characterization of mouse myelin/oligodendrocyte glycoprotein."

RT: J. Neurochem. 58:1676-1682(1992).

RL: [1]

CC: -1- FUNCTION: MINOR COMPONENT OF THE MYELIN SHEATH. MAY BE INVOLVED IN COMPLETION AND/OR MAINTENANCE OF THE MYELIN SHEATH AND IN CELL-CELL COMMUNICATION.

CC: -1- SUBUNIT: MAY FORM HOMODIMERS.

CC: -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

CC: -1- TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN THE CNS, WHERE IT IS LOCALIZED ON THE SURFACE OF MYELIN AND OLIGODENDROCYTE CYTOPLASMIC MEMBRANES.

CC: -1- DISEASE: REDUCED CONCENTRATIONS OF MOG ARE OBSERVED IN JIMPY AND DUMCKING DYSPLEINATING MOUTANT MICE.

CC: -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS ONE V-LIKE DOMAIN. BELONGS TO THE B2M/MOG SUBFAMILY.

[illegible]

FT CONFLICT 589 589 Y -> C (IN REF. 3).
 FT CONFLICT 601 601 V -> A (IN REF. 3).
 FT CONFLICT 1364 1364 E -> A (IN REF. 3).
 FT CONFLICT 2118 2119 ML -> IV (IN REF. 1).
 SQ SEQUENCE 4092 AA; 471337 MW; 309DF447E8E2D6B8 CRC64;

Query Match
 Best Local Similarity 27.3%; Pred. No. 1.5e+02; Length 4092;
 Matches 18; Conservative 13; Mismatches 33; Indels 2; Gaps 2;

QY 47 ARGDALRWITTYNTTSSKRRRRSVCHARTWCFCKYDYRRSMDTITNTIS 106
 DB 3667 ARRVYDEIMLWLYOEYVCPSTALDKKFKMIM-AMTWFCILYKFD-IESEQYKAVLTMIG 3724
 QY 107 VVSAS 112
 DB 3725 VLSSE 3730

RESULT 15
 ID MOG_RAT STANDARD: PRT; 245 AA.
 AC 063345;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE MYELIN-OLIGODENDROCYTE GLYCOPROTEIN PRECURSOR.
 GN MOG.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE; 93085763.
 RA Gardiner M.V., Amiguet P., Linington C., Mathieu J.-M.;
 RA "Myelin/oligodendrocyte glycoprotein is a unique member of the
 RA immunoglobulin superfamily";
 RA J. Neurosci. Res. 33:177-187(1992).
 RL [2]
 RN [2]
 RP SEQUENCE OF 28-245 FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE; 93376728.
 RA Pham-Dinh D., Matel M.-G., Nussbaum J.-L., Rousset G.,
 RA Pontarotti P., Roedel N., Mather I.H., Artzt K., Lindahl K.F.,
 RA "Myelin/oligodendrocyte glycoprotein is a member of the
 RA immunoglobulin superfamily encoded within the major
 RA histocompatibility complex";
 RA Proc. Natl. Acad. Sci. U.S.A. 90:7990-7994(1993).
 RN [3]
 RP STRUCTURE BY NMR OF 62-82.
 RX MEDLINE; 97354172.
 RA Albouze-Abou S., Wilson J.C., Bernard C.C.A., von Itzstein M.;
 RA "A conformational study of the human and rat encephalitogenic myelin
 RA oligodendrocyte glycoprotein peptides 35-55";
 RA Eur. J. Biochem. 246:59-70(1997).
 RL [4]
 RP FUNCTION: MINOR COMPONENT OF THE MYELIN SHEATH. MAY BE INVOLVED IN
 RP COMPLETION AND/OR MAINTENANCE OF THE MYELIN SHEATH AND IN CELL-
 RP CELL COMMUNICATION.
 CC [1] SUBUNIT: MAY FORM HOMODIMERS.
 CC [1] SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC [1] TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN THE CNS, WHERE IT IS
 CC LOCALIZED ON THE SURFACE OF MYELIN AND OLIGODENDROCYTE CYTOPLASMIC
 CC MEMBRANES.
 CC [1] DEVELOPMENTAL STAGE: A PEAK OF EXPRESSION HAS BEEN OBSERVED
 CC BETWEEN POSTNATAL DAYS 15 AND 25, COINCIDENT WITH THE PERIOD OF
 CC ACTIVE MYELINATION.
 CC [1] SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
 CC ONE V-LIKE DOMAIN. BELONGS TO THE B2M/MOG SUBFAMILY.
 CC [1] CAUTION: DO NOT CONFUSE MYELIN-OLIGODENDROCYTE GLYCOPROTEIN (MOG)
 CC WITH OLIGODENDROCYTE-MYELIN GLYCOPROTEIN (OMG).
 CC [1]

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M99485; AAA41628.1;
 DR EMBL; L21995; -; NOT_ANNOTATED_CDS.
 DR PFM; PF00047; 19; 1.
 KW Immunoglobulin domain; Glycoprotein; Myelin; Transmembrane; Signal.
 FT SIGNAL 1 27
 FT CHAIN 28 245 MYELIN-OLIGODENDROCYTE GLYCOPROTEIN.
 FT DOMAIN 28 155 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 156 176 POTENTIAL.
 FT DOMAIN 177 208 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 209 229 POTENTIAL.
 FT DOMAIN 230 245 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 44 132 IG-LIKE V-TYPE DOMAIN.
 FT DISULFD 51 125 POTENTIAL.
 FT CARBOHYD 58 58 POTENTIAL.
 SQ SEQUENCE 245 AA; 27881 MW; C97F8AD60D6A32B4 CRC64;

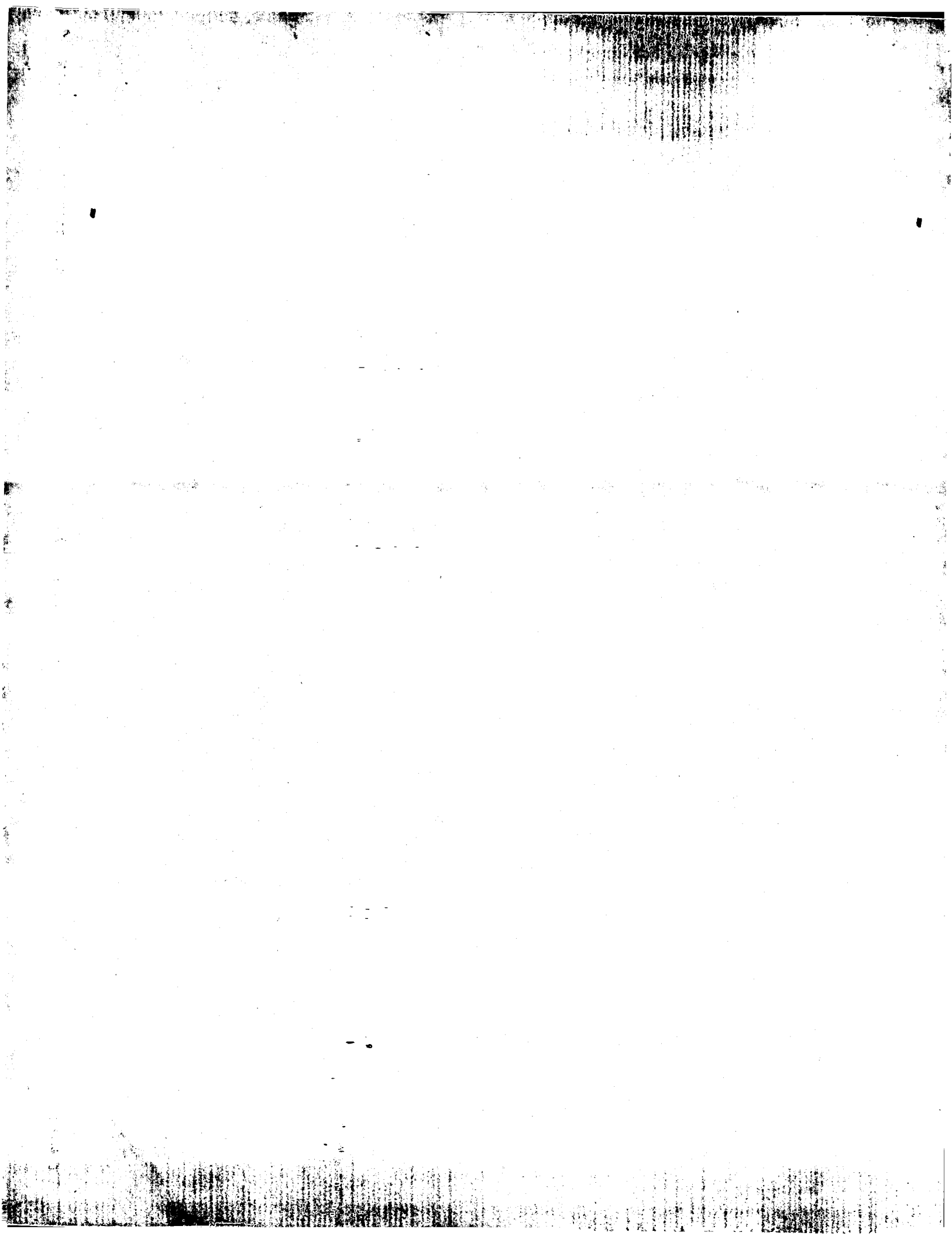
Query Match
 Best Local Similarity 10.0%; Score 62; DB 1; Length 245;
 Matches 15; Conservative 5; Mismatches 20; Indels 0; Gaps 0;

QY 19 VGMIGVAVSAYVDVYRAAHEGYIEPEARGRLDALREMI 58
 DB 64 VGMVRSPESSRVVYHYRNKGDDADQAGRYRRTLLNEST 103

Search completed: June 23, 2000, 10:14:51
 Job time: 672 sec

Mon Jun 26 09:11:01 2000

us-09-214-478-4.rsp



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2000, 10:07:44 ; Search time 61.44 Seconds
(without alignments)
128.647 Million cell updates/sec

Title: US-09-214-478-4
Perfect score: 621
Sequence: 1 MYLPLPAPPCVDSQNECVG.....SIMHDTTNTISVSAHSVQ 114

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing First 45 summaries

Database :

SPTREMBL_12: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mmc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_proteus: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	254.5	41.0	120	12	P89082
2	183	29.5	121	12	064866 human adeno
3	75	12.1	347	5	025436 mayetiola d
4	72.5	11.7	1232	1	058318 mechanococ
5	66.5	10.7	945	10	092VR3 arabidopsis
6	66	10.6	559	12	005552 ectromelia
7	65	10.5	527	11	064304 rattus leuc
8	64.5	10.4	284	2	053718 mycobacteri
9	64.5	10.4	496	10	092Q98 arabidopsis
10	64.5	10.4	1095	13	093596 brachydanio
11	64	10.3	1083	10	048839 arabidopsis
12	63.5	10.2	264	5	019843 caenorhabd
13	63.5	10.2	386	13	093370 brachydanio
14	63.5	10.2	516	3	094512 schizosacch
15	63.5	10.2	756	10	022863 arabidopsis
16	63.5	10.2	1624	2	053203 mycobacteri
17	63.5	10.2	1639	2	085019 mycobacteri
18	63	10.1	394	11	055149 rattus norv
19	63	10.1	394	11	088693 mus musculu
20	63	10.1	896	2	050747 borrelia bu

21	62.5	10.1	239	11	035258 mus musculu
22	62.5	10.1	240	11	054832 mus musculu
23	62.5	10.1	322	2	086892 streptococ
24	62.5	10.1	2843	4	096877 homo sapien
25	62	10.0	211	3	012346 saccharomye
26	62	10.0	629	2	092NP9 pseudomonas
27	62	10.0	896	1	036052 pyrococcus
28	62	10.0	969	5	021502 caenorhabd
29	62	10.0	1759	5	093187 caenorhabd
30	62	10.0	1781	5	017758 caenorhabd
31	61.5	9.9	244	5	018352 caenorhabd
32	61.5	9.9	455	10	092RM2 vigna mungo
33	61.5	9.9	771	10	022943 arabidopsis
34	61.5	9.9	886	4	095099 homo sapien
35	61.5	9.9	1413	4	075696 homo sapien
36	61.5	9.9	1661	5	077244 chlorohydra
37	61.5	9.9	2080	4	075923 homo sapien
38	61	9.8	84	6	019086 sus scrofa
39	61	9.8	215	12	080635 human timun
40	61	9.8	308	2	0939R1 sphingomona
41	61	9.8	668	12	056765 human cytom
42	60.5	9.7	210	2	P95321 mycobacteri
43	60.5	9.7	223	2	050352 lactobacill
44	60.5	9.7	898	6	062683 canis fami
45	60	9.7	151	2	007408 mycobacteri

ALIGNMENTS

RESULT 1
ID P89082 PRELIMINARY; PRT; 120 AA.
AC P89082;
DT 01-MAY-1997 (TREMREL. 03, Created)
DT 01-MAY-1997 (TREMREL. 03, Last sequence update)
DT 01-NOV-1998 (TREMREL. 08, Last annotation update)
DE ORF4.
OS Human adenovirus type 9.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96327849.
RA JAVIER R., SHENK T.;
RT "Mammary tumors induced by human adenovirus type 9: a role for the viral early region 4 gene."
RL Breast Cancer Res. Treat. 39:57-67(1996).
DR EMBL; S82508; AAB37507.1; -
SQ SEQUENCE 120 AA; 14152 MW; EE2C2416 CRC32;

Query Match 41.0%; Score 254.5; DB 12; Length 120;
Best Local Similarity 40.7%; Pred. No. 5.6e-21;
Matches 46; Conservative 26; Mismatches 40; Indels 1; Gaps 1;
DB 1 MYLPLPAPPCVDSQNECVGLGVAVSVVDVIRAAHEGYIPEAGRGDAREWYIY 60
1 MYLPLPAPPCVDSQNECVGLGVAVSVVDVIRAAHEGYIPEAGRGDAREWYIY 59
61 NYTFRSRKRRRRSVCHARTWFCFRKYDYVRSIMHDTTNTISVSAHSV 113
60 SMWTRQRKRRKRGICCSRAHFCWQYKDYKRIHINARNDSTQLAPPSI 112
RESULT 2
ID 064866 PRELIMINARY; PRT; 121 AA.
AC 064866;
DT 01-NOV-1996 (TREMREL. 01, Created)
DT 01-NOV-1996 (TREMREL. 01, Last sequence update)
DT 01-NOV-1998 (TREMREL. 08, Last annotation update)
DE ADENOVIRUS TYPE 40, COMPLETE GENOME.
OS Human adenovirus type 40.

OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 87122182.
 RA ISHINO M., SAMADA Y., YABGASHI T., DEMURA M., FUJINAGA K.;
 RT "Nucleotide sequence of the adenovirus type 40 inverted terminal
 RL repeat: close relation to that of adenovirus type 5.";
 RN Virology 156:414-416(1987).
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 88084437.
 RA VAN LOON A.E., LIGTENBERG M., REEMST A.M., SUSSENBACH J.S.,
 RT "Structure and organization of the left-terminal DNA regions of
 RL fastidious adenovirus types 40 and 41.";
 RN Gene 58:109-128(1987).
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RA ISHINO M.;
 RL Sapporo Igaku Zasshi 57:59-66(1988).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 88265890.
 RA ISHINO M., OHASHI Y., EMOTO T., SAWADA Y., FUJINAGA K.;
 RT "Characterization of adenovirus type 40 EL region.";
 RL Virology 165:95-102(1988).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 88160034.
 RA VOS H.L., VAN DER LEE F.M., REEMST A.M., VAN LOON A.E.,
 RT "The genes encoding the DNA binding protein and the 23k protease of
 RL adenovirus types 40 and 41.";
 RN Virology 163:1-10(1988).
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 89370295.
 RA KIDD A.H., ERASMUS M.J.;
 RT "Sequence characterization of the adenovirus 40 fiber gene.";
 RL Virology 172:134-144(1989).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 90111698.
 RA TOOGOOD C.I., MURALI R., BURNETT R.M., HAY R.T.;
 RT "The adenovirus type 40 hexon: sequence, predicted structure and
 RL relationship to other adenovirus hexons.";
 RN J. Gen. Virol. 70:3203-3214(1989).
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 94087748.
 RA DAVISON A.J., TELFORD E.A., WATSON M.S., MCBRIDE K., MAUTNER V.;
 RT "The DNA sequence of adenovirus type 40.";
 RL J. Mol. Biol. 234:1308-1316(1993).
 RN [9]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RA PIENIAZEK N.J., SLEWENDA S.B., PIENIAZEK D., LUFTIG R.B.;
 RT Submitted (JUN-1993) to the EMBL/Genbank/DBJ databases.
 DR EMBL: L19443; A013982.1;
 SQ SEQUENCE 121 AA; 13976 MW; 8884054F CRC32;

Query Match 29.5%; Score 183; DB 12; Length 121;
 Best Local Similarity 35.5%; Pred. No. 4.6e-13;
 Matches 38; Conservative 20; Mismatches 49; Indels 0; Gaps 0;

OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 87122182.
 RA ISHINO M., SAMADA Y., YABGASHI T., DEMURA M., FUJINAGA K.;
 RT "Nucleotide sequence of the adenovirus type 40 inverted terminal
 RL repeat: close relation to that of adenovirus type 5.";
 RN Virology 156:414-416(1987).
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 88084437.
 RA VAN LOON A.E., LIGTENBERG M., REEMST A.M., SUSSENBACH J.S.,
 RT "Structure and organization of the left-terminal DNA regions of
 RL fastidious adenovirus types 40 and 41.";
 RN Gene 58:109-128(1987).
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RA ISHINO M.;
 RL Sapporo Igaku Zasshi 57:59-66(1988).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 88265890.
 RA ISHINO M., OHASHI Y., EMOTO T., SAWADA Y., FUJINAGA K.;
 RT "Characterization of adenovirus type 40 EL region.";
 RL Virology 165:95-102(1988).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 88160034.
 RA VOS H.L., VAN DER LEE F.M., REEMST A.M., VAN LOON A.E.,
 RT "The genes encoding the DNA binding protein and the 23k protease of
 RL adenovirus types 40 and 41.";
 RN Virology 163:1-10(1988).
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 89370295.
 RA KIDD A.H., ERASMUS M.J.;
 RT "Sequence characterization of the adenovirus 40 fiber gene.";
 RL Virology 172:134-144(1989).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 90111698.
 RA TOOGOOD C.I., MURALI R., BURNETT R.M., HAY R.T.;
 RT "The adenovirus type 40 hexon: sequence, predicted structure and
 RL relationship to other adenovirus hexons.";
 RN J. Gen. Virol. 70:3203-3214(1989).
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 94087748.
 RA DAVISON A.J., TELFORD E.A., WATSON M.S., MCBRIDE K., MAUTNER V.;
 RT "The DNA sequence of adenovirus type 40.";
 RL J. Mol. Biol. 234:1308-1316(1993).
 RN [9]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RA PIENIAZEK N.J., SLEWENDA S.B., PIENIAZEK D., LUFTIG R.B.;
 RT Submitted (JUN-1993) to the EMBL/Genbank/DBJ databases.
 DR EMBL: L19443; A013982.1;
 SQ SEQUENCE 121 AA; 13976 MW; 8884054F CRC32;

Query Match 12.1%; Score 75; DB 5; Length 347;
 Best Local Similarity 26.3%; Pred. No. 1.4;
 Matches 26; Conservative 16; Mismatches 49; Indels 8; Gaps 3;

OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 87122182.
 RA ISHINO M., SAMADA Y., YABGASHI T., DEMURA M., FUJINAGA K.;
 RT "Nucleotide sequence of the adenovirus type 40 inverted terminal
 RL repeat: close relation to that of adenovirus type 5.";
 RN Virology 156:414-416(1987).
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 88084437.
 RA VAN LOON A.E., LIGTENBERG M., REEMST A.M., SUSSENBACH J.S.,
 RT "Structure and organization of the left-terminal DNA regions of
 RL fastidious adenovirus types 40 and 41.";
 RN Gene 58:109-128(1987).
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RA ISHINO M.;
 RL Sapporo Igaku Zasshi 57:59-66(1988).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 88265890.
 RA ISHINO M., OHASHI Y., EMOTO T., SAWADA Y., FUJINAGA K.;
 RT "Characterization of adenovirus type 40 EL region.";
 RL Virology 165:95-102(1988).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 88160034.
 RA VOS H.L., VAN DER LEE F.M., REEMST A.M., VAN LOON A.E.,
 RT "The genes encoding the DNA binding protein and the 23k protease of
 RL adenovirus types 40 and 41.";
 RN Virology 163:1-10(1988).
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 89370295.
 RA KIDD A.H., ERASMUS M.J.;
 RT "Sequence characterization of the adenovirus 40 fiber gene.";
 RL Virology 172:134-144(1989).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 90111698.
 RA TOOGOOD C.I., MURALI R., BURNETT R.M., HAY R.T.;
 RT "The adenovirus type 40 hexon: sequence, predicted structure and
 RL relationship to other adenovirus hexons.";
 RN J. Gen. Virol. 70:3203-3214(1989).
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RX MEDLINE: 94087748.
 RA DAVISON A.J., TELFORD E.A., WATSON M.S., MCBRIDE K., MAUTNER V.;
 RT "The DNA sequence of adenovirus type 40.";
 RL J. Mol. Biol. 234:1308-1316(1993).
 RN [9]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DUGAN;
 RA PIENIAZEK N.J., SLEWENDA S.B., PIENIAZEK D., LUFTIG R.B.;
 RT Submitted (JUN-1993) to the EMBL/Genbank/DBJ databases.
 DR EMBL: L19443; A013982.1;
 SQ SEQUENCE 121 AA; 13976 MW; 8884054F CRC32;

DR EMBL: U67534; AAB98910.1; -
 RC TYGR: M07080; -
 KW Hypothetical protein.
 SQ SEQUENCE 1232 AA; 142643 MW; FB712545 CRC32;

Query Match 11.7%; Score 72.5; DB 1; Length 1232;
 Best Local Similarity 23.8%; Pred. No. 12;
 Matches 35; Conservative 14; Mismatches 51; Indels 47; Gaps 6;

QY 3 LPALPAPVPC-----DSQECVGLVAISAV-----30
 DB 86 LPPLPTLSIHDPLEKDRVDDVKNKYKYLIG--YEGVYNNKLLLYLANFGNLNV 143
 QY 30 VDVYRAAHAGVYIEPARRLDALREMYNYNTYTERSKRDRRSYCHAFMFCRK 88
 DB 144 EYEPBPMPWGGIYKGYFETLDD---LYNLKELGRDLKPIIGVLFYRMWFAVANN 198
 QY 89 YDYVRSIWHDTTN---TISVSAH 111
 DB 199 IDYVNDLI--DIENKGAIPAVFSH 223

RESULT 5
 092VK3 PRELIMINARY; PRT; 945 AA.

AC 092VK3;
 DT 01-MAY-1999 (TREMBLREL. 10, Created)
 DT 01-MAY-1999 (TREMBLREL. 10, Last sequence update)
 DE 01-MAY-1999 (TREMBLREL. 10, Last annotation update)
 DE REVERSE-TRANSCRIPTASE-LIKE PROTEIN.
 GN T6B13.12.
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 CC core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae;
 CC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA ROUNSELEY S.D., LIN X., KAUL S., SHEA T.P., FUJII C.Y., MASON T.M.,
 RA SHEN M., RONNING C.M., FRASER C.M., SOMERVILLE C.R., VENTER J.C.;
 RA "Arabidopsis thaliana chromosome II BAC T6B13 genomic sequence."
 RL Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC005398; AAC69378.1; -
 SQ SEQUENCE 945 AA; 108084 MW; 88791777 CRC32;

Query Match 10.7%; Score 66.5; DB 10; Length 945;
 Best Local Similarity 20.7%; Pred. No. 40;
 Matches 23; Conservative 16; Mismatches 29; Indels 43; Gaps 6;

QY 15 QNECGMGLVANSVVDVYRAAHAGVY-----IEPERGRDLALR-EMYYNYT 64
 DB 180 QNKYVGW-----TQEVLVGAFMNGLYTISNGIRMPFSLREYINLVGR----- 226
 QY 65 ESKRRDRRRRSYCH-----ARTWFCFRXY-----DYVRSIM 97
 DB 226 -KGSKDRKRKRATHTHLAVLSITWCAKROSNRRORDSVKKNYERKGVW 275

RESULT 6
 005552 PRELIMINARY; PRT; 559 AA.

AC 005552;
 DT 01-NOV-1996 (TREMBLREL. 01, Created)
 DT 01-NOV-1996 (TREMBLREL. 01, Last sequence update)
 DE 01-NOV-1998 (TREMBLREL. 08, Last annotation update)
 DE PROTEIN P65 (PROTEIN B21 HOMOLOG).
 OS Ectromella virus.
 CC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
 OC Orthopoxvirus.
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-MILL. HILL;
 RA MEDLINE: 94091057.
 RA SENKEVICH T.G., MURAVNIK G.L., POZDNYAKOV S.G., CHIZHIKOV V.E.,
 RA KRAZANKINA O.I., SHCHELKUNOV S.N., KOONIN E.V., CHERNOS V.I.;
 RT "Nucleotide sequence of XhoI O fragment of ectromella virus DNA
 RT reveals significant differences from vaccinia virus."
 RL Virus Res. 30:73-88(1993).
 DR EMBL: Z14256; CAA78622.1; -
 DR PFAM: PF00651; BTF. 1.
 DR PFAM: PF01344; Kelch. 3.
 SQ SEQUENCE 559 AA; 65281 MW; 491EAE41 CRC32;

Query Match 10.6%; Score 66; DB 12; Length 559;
 Best Local Similarity 20.3%; Pred. No. 25;
 Matches 26; Conservative 26; Mismatches 42; Indels 34; Gaps 7;

QY 1 MVLPAIPVPCDSQNECVGLVANSVVDVYRAAHAGVYIEPARGRDLALREMY 60
 DB 150 MYMKNIPTLTITDAFKNAV-----FEILDII--STNDEYVYREGY-KVTILKLWLDY 200
 QY 61 NYTERS-----KRDRRRRSYCHARTW-----CFR-----KYDVRS--- 96
 DB 201 NYTEEQLCLTSCIDQINDKRSRLLYSNTTINMTSSCVKFLDNKKNRIIPQLCL 260
 QY 96 IMHDTTN 103
 DB 261 VYHDTN 268

RESULT 7
 064304 PRELIMINARY; PRT; 527 AA.

AC 064304;
 DT 01-NOV-1996 (TREMBLREL. 01, Created)
 DT 01-NOV-1996 (TREMBLREL. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBLREL. 08, Last annotation update)
 DE PROTEIN LINKED TO SYSTEM L-LIKE NEUTRAL AMINO ACID TRANSPORT
 DE ACTIVITY.
 GN LIAT.
 OS Rattus leucopus (mottle-tailed rat), and Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE: 96128073.
 RA BROEER S., BROEER A., HAMPRECHT B.;
 RT "The 4F2hc surface antigen is necessary for expression of system L-
 RT like neutral amino acid-transport activity in C6-BU-1 rat glioma
 RT cells: evidence from expression studies in Xenopus laevis oocytes."
 RL Biochem. J. 312:863-870(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SPAGUE DAWLEY, TISSUE-JETUNAL BRUSH BORDER;
 RA YAO S.Y.M., MOZIKA W.R., CHEBESMAN C.T., ELLIOTT J.F., YOUNG J.D.;
 RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN-SPAGUE DAWLEY, TISSUE-LIVER;
 RX MEDLINE: 98395066.
 RA KANAI Y., SEGAWA H., MIYAMOTO K., UCHINO H., TAKEDA E., ENDO H.;
 RT "Expression cloning and characterization of a transporter for large
 RT neutral amino acids activated by the heavy chain of 4F2 antigen
 RT (CD98)."
 RL J. Biol. Chem. 273:23629-23632(1998).
 DR EMBL: X89325; CAA61509.1; -
 DR EMBL: U59324; AAC53560.1; -
 DR EMBL: AB015433; BAA33036.1; -
 DR PFAM: PF00128; alpha-amylase. 1.
 SQ SEQUENCE 527 AA; 58072 MW; 6C2869F1 CRC32;

DT 01-JUN-1998 (Tremblrel. 06, Created)
 DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE PTATIVE UBIQUITIN SPECIFIC PROTEASE.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN:CV. COLUMBIA.
 RA ROUSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,
 RA SKERIS S.M., KADU S., MASON T.M., KERLAUGE A.R., ADAMS M.D.,
 RA SKEVERVILLE C.R., VENTER J.C.
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC003974; AAC04485.1;
 DR PRAM: PF00442; UCH-1; 1.
 DR Protease.
 KW SEQUENCE 1083 AA; 120769 MW; C61CA539 CRC32;
 SQ

Query Match 10.3%; Score 64; DB 10; Length 1083;
 Best Local Similarity 24.3%; Pred. No. 88;
 Matches 18; Conservative 12; Mismatches 30; Indels 14; Gaps 1;

QY 14 SQNECVMLGVAYSVVDVIRAAAEVYI-----EPARGRLALREMIY 59
 ID 019843 PRELIMINARY; PRT; 264 AA.
 AC 019843;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
 DE SIMILARITY TO HUMAN FOLICULAR VARIANT TRANSLOCATION PROTEIN 1
 DE PRECURSOR.
 GN F27D9.6.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdillia; Rhabdillia;
 OC Rhabdillia; Rhabdillidae; Rhabdillidae; Peloderinae; Caenorhabdillia.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94150718.
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BODTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
 RA CRAYON M., DEAR S., DU Z., DUBBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HARKINS T., HILLIER L., JIE M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATEILLE P.,
 RA LIGHTING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULLSTON J.,
 RA THIERY-MIEG J., THOMAS K., THOMAS K., VAUGHAN K., WATTESTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOLDMAN P.,
 RT 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA BENTLEY D.,
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA WATERSTON R.,
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL: U49829; AAA93385.1;
 SQ SEQUENCE 264 AA; 29808 MW; 90319321 CRC32;
 Query Match 10.2%; Score 63.5; DB 5; Length 264;
 Best Local Similarity 23.2%; Pred. No. 19;
 Matches 26; Conservative 17; Mismatches 48; Indels 21; Gaps 6;

QY 13 DSQNECV---GW---LGVAYSVVDVIRAAAEVYIEPARGRLALREMIYNYT 64
 ID 093370 PRELIMINARY; PRT; 386 AA.
 AC 093370;
 DT 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, Last annotation update)
 DE TRANSCRIPTION FACTOR.
 GN PAX2.2.
 OS Brachydanio rerio (Zebrafish) (Zebra danio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
 OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Cyprinidae; Rasbora; Danio.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 98337826.
 RA PEEFER P.L., GERSTYER T., LUN K., BRAND M., BUSSLINGER M.,
 RT "Characterization of three novel members of the zebrafish Pax2/5/8
 RT family: dependency of Pax and Pax8 expression on the Pax2.1 (not)
 RT function."
 RL Development 125:3063-3074(1998).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.
 DR EMBL: AF072547; AAC31811.1;
 DR HSP: P06601; 1PDN.
 DR PROSITE: PS00034; PAIRED_BOX; 1.
 DR PRAM: PF00292; PAX; 1.
 DR PRINTS: PR00027; PAIREDBOX.
 KW Paired box; DNA-binding; Developmental protein; Nuclear protein.
 SQ SEQUENCE 386 AA; 41541 MW; 5BCDF72B CRC32;
 Query Match 10.2%; Score 63.5; DB 13; Length 386;
 Best Local Similarity 21.6%; Pred. No. 30;
 Matches 16; Conservative 22; Mismatches 25; Indels 11; Gaps 2;

QY 2 VLPLAPAPVDSQNECVMLGVAYSVVDVIR-----AAHGVYIEPARGRLD 52
 ID 094512 PRELIMINARY; PRT; 516 AA.
 AC 094512;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
 DE PROBABLE INVOLVEMENT IN ERGOSTEROL BIOSYNTHESIS.
 GN SPC646.08C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;

OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972H-;
 RA SEEGER K., HARRIS D., WOOD V., RAJANDREAM M.A., BARRELL B.G.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL035216; CAA22812.1; -
 SQ SEQUENCE 516 AA; 58692 MW; BA471658 CRC32;

Query Match 10.2%; Score 63.5; DB 3; Length 516;
 Best Local Similarity 22.9%; Pred. No. 42;
 Matches 25; Conservative 14; Mismatches 47; Indels 23; Gaps 4;

QY 21 WGVAVASAV-----VDVRAAHGEGYIEPEARGRLDLREWIYNYNTERS 67
 DB 354 WIGKPSRLVGVGYRYPDSNDTDSIKAVPKENILATFKGN-----WRNCIFYSTAGESE 408
 QY 68 KRR--DRRRRSYCHARTWFCFRKYDYVRSIMWDTTNTISVSAHSVQ 114
 DB 409 SRMLVDLNLDLVHKRCPLDKQFPESRKIMFPVTHN--ILAKHYTQ 454

RESULT 15

ID 022863 PRELIMINARY; PRT; 756 AA.
 AC 022863;
 DT 01-JAN-1998 (TREMblrel. 05, Created)
 DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
 DE HYPOTHETICAL PROTEIN.
 GN T01024.23.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA ROUSLEY S.D., TSCHDY M.M., LIN X., KETCHUM K.A., CROSBY M.L.,
 RA BRANDON R.C., SPRIGGS T.A., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,
 RA SOMERVILLE C.R., VENTER J.C.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC002335; AAB64317.1; -
 DR PFAM; PF00566; TBC; 1
 SQ SEQUENCE 756 AA; 85182 MW; F71F6CF4 CRC32;

Query Match 10.2%; Score 63.5; DB 10; Length 756;
 Best Local Similarity 34.1%; Pred. No. 66;
 Matches 14; Conservative 7; Mismatches 17; Indels 3; Gaps 1;

QY 21 WGVAVASAVVIVRAAHGEGYIEPEARGRLD--ALREWI 58
 DB 355 WLMTLRIYVVDVVRTDSHLEFYEDPGLGRMSDILAVYAVV 395

Search completed: June 23, 2000, 10:07:48
 Job time: 951 sec

Mon Jun 26 09:10:50 2000

us-09-214-478-3.rml

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 23, 2000, 09:51:07 ; Search time 42.97 Seconds
(without alignments)
1043.631 Million cell updates/sec

Title: US-09-214-478-3

Sequence: 1 atgtctctccagctctcc.....cggcgactcgcgtagag 345

Scoring table: GAPOP 10.0, Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

Issued Patents NA.*
1: /cgn2_6/ptodata/1/lna/5/COMB.seq.*
2: /cgn2_6/ptodata/1/lna/5/COMB.seq.*
3: /cgn2_6/ptodata/1/lna/5/COMB.seq.*
4: /cgn2_6/ptodata/1/lna/5/COMB.seq.*
5: /cgn2_6/ptodata/1/lna/6/COMB.seq.*
6: /cgn2_6/ptodata/1/lna/6/COMB.seq.*
7: /cgn2_6/ptodata/1/lna/6/COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query length	ID	Description
1	345	100.0	US-08-735-609-4	Sequence 4, Appl 1
2	345	100.0	US-08-374-483-6	Sequence 6, Appl 1
3	345	100.0	US-08-735-609-1	Sequence 1, Appl 1
4	345	100.0	US-08-379-452-43	Sequence 43, Appl 1
5	80	23.2	US-08-462-014-2	Sequence 2, Appl 1
6	54	15.7	US-08-752-760A-1	Sequence 1, Appl 1
7	29	8.4	US-08-452-267-2	Sequence 2, Appl 1
8	29	8.4	US-08-723-624-18	Sequence 18, Appl 1
9	29	8.4	US-08-630-820-5	Sequence 3, Appl 1
10	29	8.4	US-08-723-624-19	Sequence 19, Appl 1
11	29	8.4	US-08-452-267-3	Sequence 3, Appl 1
12	29	8.4	US-08-318-72A-2	Sequence 2, Appl 1
13	29	8.4	US-08-882-704A-4	Sequence 4, Appl 1
14	29	8.4	US-08-819-866-1	Sequence 1, Appl 1
15	29	8.4	US-09-023-715-1	Sequence 1, Appl 1
16	29	8.4	US-08-819-866-2	Sequence 2, Appl 1
17	29	8.4	US-09-023-715-2	Sequence 2, Appl 1
18	29	8.4	US-08-453-924-4	Sequence 4, Appl 1
19	29	8.0	US-08-471-791-28	Sequence 28, Appl 1
20	29	8.0	PCT-US91-01746-28	Sequence 28, Appl 1
21	29	7.9	5432081-1	Sequence 1, Appl 1
22	29	7.9	US-08-525-507-14	Sequence 14, Appl 1
23	29	7.9	US-08-304-309-3	Sequence 3, Appl 1
24	29	7.9	US-08-991-942-3	Sequence 3, Appl 1
25	29	7.9	US-08-373-652-9	Sequence 9, Appl 1
26	29	7.9	PCT-US95-16311-9	Sequence 9, Appl 1
27	29	7.8	US-08-605-106-6	Sequence 6, Appl 1

ALIGNMENTS

28	26.6	7.7	2740	1	US-08-100-247-4	Sequence 4, Appl 1
29	26.6	7.7	2740	2	US-08-483-146A-4	Sequence 4, Appl 1
30	26.6	7.7	2740	2	US-08-232-513A-5	Sequence 5, Appl 1
31	26.6	7.7	2740	2	US-08-484-594A-4	Sequence 4, Appl 1
32	26.4	7.7	2380	7	5268463-1	Sequence 7, Appl 1
33	26.4	7.7	7218	1	US-08-232-463-14	Sequence 14, Appl 1
34	26.2	7.6	34	3	US-08-572-128-2	Sequence 2, Appl 1
35	26.2	7.6	34	4	US-08-757-023A-2	Sequence 2, Appl 1
36	26.2	7.6	1307	4	US-08-960-022-17	Sequence 17, Appl 1
37	26.2	7.6	1849	4	US-08-676-166A-1	Sequence 1, Appl 1
38	26	7.5	852	4	US-08-751-233A-5	Sequence 5, Appl 1
39	25.8	7.5	1320	5	US-08-718-904-76	Sequence 76, Appl 1
40	25.8	7.5	3181	2	US-08-653-086-1	Sequence 1, Appl 1
41	25.6	7.4	295	5	US-08-765-332-167	Sequence 167, Appl 1
42	25.6	7.4	1897	1	US-08-245-688-3	Sequence 3, Appl 1
43	25.6	7.4	1897	1	US-08-245-688-3	Sequence 3, Appl 1
44	25.6	7.4	1897	1	US-08-245-688-5	Sequence 5, Appl 1
45	25.6	7.4	1897	1	US-08-245-688-7	Sequence 7, Appl 1

RESULT 1
US-08-735-609-4/c

Sequence 4, Application US/08735609
Patent No. 595360

GENERAL INFORMATION:

APPLICANT: Chamberlain, Jeffrey S.

APPLICANT: Amalfitano, Andrea

APPLICANT: Hauser, Michael A.

APPLICANT: Kumar-Singh, Rajendra

APPLICANT: Hartigan-O'Connor, Dennis J.

TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSER: Medlen & Carroll, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/735,609

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Ingolia, Diane E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: UM-02484

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 34303 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA"

Query Match 100.0%; Score 345; DB 4; Length 34303;
Best Local Similarity 100.0%; Pred. No. 2.3e-112;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggttctccacgctctccgctctccgctgtgtgactcgacagcaaatgttagt 60
|||||
Db 32711 ATGGTCTTCACAGCTCTCCGCTCCCTCCGCTGTGACTCGCAGAACGAATGTAGCT 32652
QY 61 tggctgggtgtggtctattctcgtgtgtgtatcaaggcagcggtcgaagaa 120
|||||
Db 32651 TGGCTGGGTGTGCTTATCTCGGTGTGTGATGATACAGGGCAGCGGCGCATGAAGG 32592
QY 121 gtttaccatgaacccggaagcagggggcgctgtgatcttggagaagtgatatactac 180
|||||
Db 32591 GTTTACATAGAACCCGAGCCAGGGGGCGCTGTGATGATGAGATATATCTAC 32532
QY 181 aactactacagagcgatcttaagcggcgagacggagacgacatcttctgacgac 240
|||||
Db 32531 AACTACTACAGAGCGACTTAAGCGCGGAGACGGAGACGATCTGTTGTACGCC 32472
QY 241 cgcacctgttctgtctcaggaataatgactacgtccggttcacattggcatgacct 300
|||||
Db 32471 CGCACTGTGTTGCTTCAGGAATATGACTGTCGGCGTTCATTTGGCATGACACT 32412
QY 301 acgaccacacgactcgtgtgtctcggcgacatcgtaacgtag 345
|||||
Db 32411 ACGACCAACAGATCTCGGTGTGTCTCGGCGCACTCCGTACAGTAG 32367

RESULT 2
US-08-374-483-6/c
Sequence 6, Application US/08374483
Patent No. 5880102
GENERAL INFORMATION:
APPLICANT: GEORGE, SAMUEL E.
APPLICANT: BLAZING, MICHAEL A.
TITLE OF INVENTION: ADENOVIRAL VECTOR SYSTEM
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON & VANDERHAYE P. C.
STREET: 1100 NORTH GLEBE ROAD, 8TH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/374,483
FILING DATE: 17-JAN-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: WILSON, MARY J.
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 1579-83
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4100
TELEFAX: (703) 816-4100
TELEX: 200797 NIXN UR
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 34382 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-374-483-6

Query Match 100.0%; Score 345; DB 3; Length 34382;
Best Local Similarity 100.0%; Pred. No. 2.3e-112;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggttctccacgctctccgctctccgctgtgtgactcgacagcaaatgttagt 60

Db 32543 ATGGTCTTCACAGCTCTCCGCTCCCTCCGCTGTGACTCGCAGAACGAATGTAGCT 32484
QY 61 tggctgggtgtggtctattctcgtgtgtgtatcaaggcagcggtcgaagaa 120
|||||
Db 32483 TGGCTGGGTGTGCTTATCTCGGTGTGTGATGATACAGGGCAGCGGCGCATGAAGG 32474
QY 121 gtttaccatgaacccggaagcagggggcgctgtgatcttggagaagtgatatactac 180
|||||
Db 32423 GTTTACATAGAACCCGAGCCAGGGGGCGCTGTGATGATGAGATATATCTAC 32364
QY 181 aactactacagagcgatcttaagcggcgagacggagacgacatcttctgacgac 240
|||||
Db 32363 AACTACTACAGAGCGACTTAAGCGCGGAGACGGAGACGATCTGTTGTACGCC 32304
QY 241 cgcacctgttctgtctcaggaataatgactacgtccggttcacattggcatgacct 300
|||||
Db 32303 CGCACTGTGTTGCTTCAGGAATATGACTACGTCGGCGTTCATTTGGCATGACACT 32244
QY 301 acgaccacacgactcgtgtgtctcggcgacatcgtaacgtag 345
|||||
Db 32243 ACGACCAACAGATCTCGGTGTGTCTCGGCGCACTCCGTACAGTAG 32199

RESULT 3
US-08-735-609-1/c
Sequence 1, Application US/08735609
Patent No. 5953560
GENERAL INFORMATION:
APPLICANT: Chamberlain, Jeffrey S.
APPLICANT: Amalfitano, Andrea
APPLICANT: Hauser, Michael A.
APPLICANT: Kumar-Singh, Rajendra
APPLICANT: Hartigan-O'Connor, Dennis J.
TITLE OF INVENTION: IMPROVED ADENOVIRUS VECTORS
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/735,609
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: UM-02484
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 35935 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA"
US-08-735-609-1

Query Match 100.0%; Score 345; DB 4; Length 35935;
Best Local Similarity 100.0%; Pred. No. 2.3e-112;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggtctccagctctccgcctcccggtgtgtgacccgagacgaatgttagt 60
|||||
Db 34342 ATGGTCTTCCAGCTCTTCCCGTCTCCGTTGTGACCTCGAAGACGAATGTAGT 34283
QY 61 tggctggtgtgtgtatctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 120
|||||
Db 34282 TGCGTGGGTGTGCTTATCTCGGGTGTGATGTATACAGGCGCGCATGAAGA 34223
QY 121 gttacatagaacccgagacgaagggcgccgtgtgtgtgtgtgtgtgtgtgtgtgt 180
|||||
Db 34222 GTTACATAGAACCCGAGACCGGCGCCCTGGATGCTTGAAGAGTGAATATAC 34163
QY 181 aactacacacagagagatctaaagcgcgagacggagacgagatctgtgtcacg 240
|||||
Db 34162 AACTACTACACAGAGCATTTAAGCGCGAGACCGAGACGAGATCTGTTGTACGCC 34103
QY 241 cgcacctgt 300
|||||
Db 34102 CGCAGCTGTTGTGTCTTCAAGAAATATGACTACGTCGCGGCTTCCATTGGCATGACACT 34043
QY 301 acgacacacacgactcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 345
|||||
Db 34042 ACGACCAACAGATCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 33998

RESULT 4
US-08-379-452-43/c
Sequence 43, Application US/08379452

Patent No. 6040174
GENERAL INFORMATION:
APPLICANT: IMLER, Jean-Luc
APPLICANT: MEHTALI, Majid
APPLICANT: PAVIRANI, Andrea
TITLE OF INVENTION: DEFECTIVE ADENOVIRUSES AND CORRESPONDING
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: 1737 King Street, Suite 500
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22314-2756
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/379,452
FILING DATE: 26-JAN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR94/00624
FILING DATE: 27-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 93 06482
FILING DATE: 28-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Dadio, Susan M.
REGISTRATION NUMBER: 40,373
REFERENCE/DOCKET NUMBER: 029395-002
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 35935 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-379-452-43

Query Match 100.0%; Score 345; DB 5; Length 35935;
Best Local Similarity 100.0%; Pred. No. 2.3e-112;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atggtctccagctctccgcctcccggtgtgtgacccgagacgaatgttagt 60
|||||
Db 34342 ATGGTCTTCCAGCTCTTCCCGTCTCCGTTGTGACCTCGAAGACGAATGTAGT 34283
QY 61 tggctggtgtgtgtatctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 120
|||||
Db 34282 TGCGTGGGTGTGCTTATCTCGGGTGTGATGTATACAGGCGCGCATGAAGA 34223
QY 121 gttacatagaacccgagacgaagggcgccgtgtgtgtgtgtgtgtgtgtgtgtgt 180
|||||
Db 34222 GTTACATAGAACCCGAGACCGGCGCCCTGGATGCTTGAAGAGTGAATATAC 34163
QY 181 aactacacacagagagatctaaagcgcgagacggagacgagatctgtgtcacg 240
|||||
Db 34162 AACTACTACACAGAGCATTTAAGCGCGAGACCGAGACGAGATCTGTTGTACGCC 34103
QY 241 cgcacctgt 300
|||||
Db 34102 CGCAGCTGTTGTGTCTTCAAGAAATATGACTACGTCGCGGCTTCCATTGGCATGACACT 34043
QY 301 acgacacacacgactcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 345
|||||
Db 34042 ACGACCAACAGATCTGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 33998

RESULT 5
US-08-462-014-2
Sequence 2, Application US/08462014

Patent No. 5756283
GENERAL INFORMATION:
APPLICANT: Wilson, James M.
APPLICANT: Fisher, Krishna J.
APPLICANT: Geo, Guang-ping
TITLE OF INVENTION: Method for Improved Production of
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr., P.O. Box 457
CITY: Spring House
STATE: PA
COUNTRY: USA
ZIP: 19477
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,014
FILING DATE: 05-JUN-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: UPNH1337USA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 8299 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
US-08-462-014-2

Query Match 23.2%; Score 80; DB 2; Length 8299;
Best Local Similarity 100.0%; Pred. No. 1.8e-18;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 266 atgactgctccgcgcgttcatttggtgcatgacactgacacacagatctggttctc 335
|||||
DB 3402 ATGCTACGTCCGGCGGTTCATTGGCATGACACTGACCAACACGATCTCGTTCTC 3461

QY 326 cggcgacactccgtacagtag 345
|||||
DB 3462 CGGGGACACTCCGTACAGTAG 3481

RESULT 6
US-08-752-760A-1/C
Sequence 1, Application US/08752760A
Patent No. 5877011

GENERAL INFORMATION:
APPLICANT: Armentano, Donna
APPLICANT: Gregory, Richard J.
APPLICANT: Smith, Alan E.
TITLE OF INVENTION: CHIMERIC ADENOVIRAL VECTORS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSER: Baker & Botts, L.L.P.
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: NY
COUNTRY: U.S.A.
ZIP: 10112

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/752,760A
FILING DATE: 20-NOV-1996
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:
NAME: Seide, Rochelle K
REGISTRATION NUMBER: 32,300
REFERENCE/DOCKET NUMBER: A31385
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-705-5000
TELEFAX: 212-705-5020
TELEX:

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 35081 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-752-760A-1

Query Match 15.7%; Score 54; DB 3; Length 35081;
Best Local Similarity 52.2%; Pred. No. 7e-09;
Matches 145; Conservative 0; Mismatches 130; Indels 3; Gaps 1;

QY 1 atggtcttcacgctctccgcgtctcccgctgctgacgcgacgaagtgtgtag 60
|||||
DB 33460 ATGGTCTTCACATCCATCCACCGCCCTCTGATGATAGAC--AAGCAGCATTTAAC 33404

QY 61 tggcgaggtgtggttattctgcgtgtgtatcagggcgcgccatgaaga 120
|||||
DB 33403 TGGATGGGAGATGGCTTAAGAGATCTGCTGATGATGAGGGGGAATTCGATGGACGG 33344

QY 121 gttacatagaaccggaagcagggggcgctgagcttctgagagagtgatatactac 180
|||||

DB 33343 TTTTATTTCATCAGATGCAGAGAGACTTCTCCAGAACTTCGAGATGATGACTTC 33284

QY 181 aactactacagagcgatctaaagcgagacggagacagatctgtgtacgcc 240
|||||

DB 33283 AGTTGATGACTGACGAGCGCAGAGGAAAGACGGCCGAGAGGGGATATATGCTGTCA 33224
QY 241 cgaactgtgttgcgttcaggaatatgactgcgcc 278
|||||

DB 33223 AGGGCCATTCTGCTGGCAGAGATGACAGAGTAGC 33186

RESULT 7
US-08-452-267-2
Sequence 2, Application US/08452267
Patent No. 5801027

GENERAL INFORMATION:
APPLICANT: Bennett, Malcolm
APPLICANT: May, Sean
APPLICANT: Ramsay, Nicholas
TITLE OF INVENTION: Control of Genes in Transgenic Plants
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSER: Barnes & Thornburg
STREET: 11 South Meridian
CITY: Indianapolis
STATE: IN
COUNTRY: USA
ZIP: 46204

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,267
FILING DATE: 26-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Breen, John P
REGISTRATION NUMBER: 38,833
REFERENCE/DOCKET NUMBER: 6653-25744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 231-7745
TELEFAX: (317) 231-7433

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2633 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: pMIGIT(insert)
US-08-452-267-2

Query Match 8.4%; Score 29; DB 2; Length 2633;
Best Local Similarity 50.4%; Pred. No. 1.3;
Matches 71; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 23 ctctcccggtgtgacacgcagacgaagtgtgagtggtggtgtgtgtattctg 82
|||||

DB 1071 CGCATGTCGGGCAAGATGATACACCCCTGTGACTGCGCAGGTGGTGGCCATGGTGA 1130
QY 83 cgggtgtgtatgtatcaagggcgagcgatgaagagttacatagaaccggaagcca 142
|||||

DB 1131 ATGTCAACGTTTGAACGTCGATGCGGATCAACAGTGGTTGCACTGGCAAGGACATA 1190
QY 143 gggggcgccgtgattgcttga 163
|||||
DB 1191 GCGGGACTTTGCAAGTGGTGA 1211

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ENHANCING
THE EXPRESSION OF GENES IN PLANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESSES:
ADDRESSEE: ARNOLD, WHITE & DUKREE
STREET: P.O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/723,624
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: BTIP:002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 3824 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-723-624-19

Query Match 8.4%; Score 29; DB 3; Length 3824;
Best Local Similarity 50.4%; Pred. No. 1.6;
Matches 71; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 23 cctctccgctgtgtactgcagagaagaatgttaggtgctgggtggtcttattctg 82
DB 2584 CGCATGTGCGCGGCAAGACTGTACACCGCGCTGTGACTGAGTGCGCAAGGCG 2643

QY 83 cgtgtgtgagtattatcagggcagcgcgatgaagagtttcatagaccggaagcca 142
DB 2644 ATGTCAAGCGTTGACGTGCGGAGTCCGATCACAGGTGTTGCAACTGACAGGCACTA 2703

QY 143 gggggcgccctgagctttga 163
DB 2704 GCGGGACTTTGCAAGTGCTGA 2724

RESULT 11
US-08-452-267-3
Sequence 3, Application US/08452267
Patent No. 5801027
GENERAL INFORMATION:
APPLICANT: Bennett, Malcolm
APPLICANT: May, Sean
TITLE OF INVENTION: Control of Genes in Transgenic Plants
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Barnes & Thornburg
STREET: 11 South Meridian
CITY: Indianapolis
STATE: IN
COUNTRY: USA
ZIP: 46204
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,267
FILING DATE: 26-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Breen, John P.
REGISTRATION NUMBER: 38,833
REFERENCE/DOCKET NUMBER: 6653-25744
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 231-7745
TELEFAX: (317) 231-7433
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 5534 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
CLONE: pUMIGIT
US-08-452-267-3

Query Match 8.4%; Score 29; DB 2; Length 5534;
Best Local Similarity 50.4%; Pred. No. 2;
Matches 71; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 23 cctctccgctgtgtactgcagagaagaatgttaggtgctgggtggtcttattctg 82
DB 1066 CGCATGTGCGCGGCAAGACTGTACACCGCGCTGTGACTGAGTGCGCAATGCTG 1125

QY 83 cgtgtgtgagtattatcagggcagcgcgatgaagagtttcatagaccggaagcca 142
DB 1126 ATGTCAAGCGTTGACGTGCGGAGTCCGATCACAGGTGTTGCAACTGACAGGCACTA 1185

QY 143 gggggcgccctgagctttga 163
DB 1186 GCGGGACTTTGCAAGTGCTGA 1206

RESULT 12
US-08-318-772A-2
Sequence 2, Application US/08318772A
Patent No. 5679558
GENERAL INFORMATION:
APPLICANT: Gobel, Elke
APPLICANT: NAKAKIDO, Fumio
TITLE OF INVENTION: Transformation of monocot cells
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22046
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC/AT
OPERATING SYSTEM: DOS version 3.3
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/318,772A
FILING DATE: 15-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP93/00905
FILING DATE: 14-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 92401066.3
FILING DATE: 15-APR-1992

MOLECULE TYPE: DNA (genomic)
US-08-819-866-1

Query Match 8.4% Score 29; DB 3; Length 14683;
Best Local Similarity 50.4%; Pred. No. 3.4;
Matches 71; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 23 ctctcccggtgtgtactgcagacgaatgtgtagttggctgggtgtgtattctg 82
DB 7569 CGCATGTGCGGCGCAAGACTGTAAACACGCGCTGTGTGACTGCGAGGTGGTCCATGTGTG 7510
QY 83 cgtgtgtgtatgtatcagggcagcgccgcatgaaggattacatagaacccgaagcca 142
DB 7509 ATGTCAAGCTTGAACCTGATGCGGATCAACAGGTGTGCAACTGACCAAGGCACTA 7450
QY 143 gggggcgccctgagatgttga 163
DB 7449 GCGGGACTTTGCCAAGTGTGA 7429

RESULT 15

US-09-023-715-1/c
Sequence 1, Application US/09023715
Patent No. 5998144

GENERAL INFORMATION:

APPLICANT: REEF, Mitchell E.
APPLICANT: BARRETT, Richard Spence
APPLICANT: MCLAUGHLIN, Karen Rella
TITLE OF INVENTION: NOVEL METHOD FOR INTEGRATING GENES AT
TITLE OF INVENTION: SPECIFIC SITES IN MAMMALIAN CELLS VIA HOMOLOGOUS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentInRelease #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,715
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/819,866
FILING DATE: 14-MAR-1997

ATTORNEY/AGENT INFORMATION:

NAME: Teskin, Robin L.

REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 012712-352

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-2021
TELEFAX: (703) 836-6620

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 14683 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-09-023-715-1

Query Match 8.4% Score 29; DB 4; Length 14683;
Best Local Similarity 50.4%; Pred. No. 3.4;
Matches 71; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 23 ctctcccggtgtgtactgcagacgaatgtgtagttggctgggtgtgtattctg 82
DB 7569 CGCATGTGCGGCGCAAGACTGTAAACACGCGCTGTGTGACTGCGAGGTGGTCCATGTGTG 7510
QY 83 cgtgtgtgtatgtatcagggcagcgccgcatgaaggattacatagaacccgaagcca 142
DB 7509 ATGTCAAGCTTGAACCTGATGCGGATCAACAGGTGTGCAACTGACCAAGGCACTA 7450
QY 143 gggggcgccctgagatgttga 163
DB 7449 GCGGGACTTTGCCAAGTGTGA 7429

Search completed: June 23, 2000, 09:51:55
JOB time: 6701 sec

Mon Jun 26 09:10:50 2000

us-09-214-478-3.rni

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2000, 10:03:34 ; Search time 20.4 Seconds

(Without alignments)
132.364 Million cell updates/sec

Title: US-09-214-478-4

Sequence: 1 MVLPAAPVCDSDQNECVG.....SIWHDITNTISVSAHSVQ 114

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: A_Geneseq_36.1*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	11.0	523	1 P80674	NTRA gene product.
2	63	10.1	394	1 W09088	Human ceramide glu
3	63	10.1	394	1 Y05381	Human HCV repress
4	62.5	10.1	2594	1 W14748	IgG-Fc binding pro
5	62.5	10.1	5405	1 W14749	IgG-Fc binding pro
6	62	10.0	150	1 W97817	Rat myelin oligode
7	61.5	9.9	1661	1 W76041	Hydra head activat
8	61.5	9.9	2763	1 W10344	Maize dwarf mosaic
9	60	9.7	118	1 R22421	Antibody produced
10	60	9.7	118	1 R54758	Humanised HmFgl he
11	60	9.7	133	1 Y13128	Human secreted pro
12	60	9.7	323	1 W56767	Homo sapiens PSP1
13	60	9.7	377	1 W56774	Homo sapiens PSP1
14	60	9.7	413	1 R13506	P. denitrificans CO
15	60	9.7	436	1 W56776	Homo sapiens PSP1
16	60	9.7	458	1 W56772	Homo sapiens PSP1
17	60	9.7	458	1 W56773	Homo sapiens PSP1
18	60	9.7	458	1 W56777	Homo sapiens PSP1
19	60	9.7	458	1 W56777	Homo sapiens PSP1
20	60	9.7	461	1 W89558	Triticum sp. cyste
21	60	9.7	529	1 W73473	Human cancer-relat
22	59.5	9.6	343	1 W21768	Protein encoded by
23	59.5	9.6	343	1 W21769	Mouse prolamine 1
24	59	9.5	176	1 W33695	Human ceramide glu
25	59	9.5	423	1 W56768	Human ceramide glu
26	59	9.5	471	1 W61216	Streptococcus pneu
27	59	9.5	649	1 R15630	Bt109p insecticid
28	58.5	9.4	117	1 R53924	Heavy chain variab
29	58.5	9.4	117	1 R53924	Heavy chain variab
30	58	9.3	215	1 R04026	O gene product of
31	57.5	9.3	177	1 R53925	Heavy chain variab
32	57.5	9.3	177	1 R53925	Kozjak mosaic viru
33	57	9.2	135	1 W09814	Reb-VH3418 fusion
34	57	9.2	215	1 W13058	HIV-2 provirus-enc

35	57	9.2	264	1 W35562	HindIII-EcoRI inse
36	57	9.2	274	1 W09819	VH4715-Linker-VH34
37	57	9.2	521	1 W74802	Human secreted pro
38	57	9.2	532	1 W35565	Human secreted pro
39	57	9.2	1019	1 W43394	Singapore horseho
40	57	9.2	1019	1 W94302	Horseho crab fac
41	57	9.2	15281	1 R44929	T. niyeum Cyclospo
42	56.5	9.1	115	1 R54813	SPA-reactive VH re
43	56.5	9.1	117	1 R53929	Heavy chain variab
44	56.5	9.1	3054	1 R40841	Translation of REV
45	56	9.0	120	1 R13657	Murine OKTA heavy

ALIGNMENTS

RESULT 1	
ID P80674	Standard; protein; 523 AA.
AC P80674	
DT 24-OCT-1990 (first entry)	
DE NTRA gene product.	
KW NTRA, sigma factor; transcriptional activation; nitrogen assimilation;	
KW fixation; nit operon; C4-dicarboxylate transport.	
OS Rhizobium meliloti.	
PN EP-292984-A.	
PD 30-NOV-1988.	
PE 27-MAY-1988; 108482.	
PR 29-MAY-1987; US-055228.	
PA (GENO-) Gen Hospital Corp.	
PI Ronson C, Ausubel F.	
DR WPI: 88-339561/48.	
DR N-PSDB: N81266.	
PT Cloned rhizobium meliloti ntra gene - for use in C4-dicarboxylate	
PT transport, nitrate assimilation, symbiotic nitrogen fixation and	
PT identifying pathogenic genes.	
PS Claim 1; Page 12; 19pp; English.	
CC The ntra gene product is a transcriptional activator. It controls the	
CC processes of nitrate assimilation, symbiotic nitrogen fixation and	
CC C4-dicarboxylate transport, working in conjunction with NtrC, NifA and	
CC DCD resp. It can also be used to identify genes involved in	
CC pathogenesis.	
CC See also P82126.	
CC Sequence 523 AA;	
QY	Query Match 11.0%; Score 68; DB 1; Length 523;
DB	Best local similarity 28.4%; Pred. No. 5.2; Mismatches 36; Indels 12; Gaps 4;
DB	Matches 27; Conservative 20;
QY	23 GVAASAVVDVIRAAHAG--VTEPEARGRDLALREWIYYNTERSKRDRRRSVCH 79
DB	221 GVEALITDYYVRAAPDGGWLELNPDLPRV-----LVNHDFYEIS--RSSRNSGEQ 343
QY	80 ARTWCFKRYDVYRSIWDITNTISVSAHSVQ 114
DB	344 AFLNECLONAMLTRSL--DQFARTIMKVASIEIV 376
RESULT 2	
ID W09088	Standard; Protein; 394 AA.
AC W09088	
DT 03-JUL-1997 (first entry)	
DE Human ceramide glucosyltransferase.	
KW Ceramide glucosyltransferase; GlcT-1; glucosylglycolipid biosynthesis;	
KW glycosylation; glucosylphosphatidyl; glucosylceramide; human.	
OS Homo sapiens.	
PN EP-748868-A2.	
PD 18-DEC-1996.	
PF 13-JUN-1996; 109447.	
PR 15-JUN-1995; JP-148472.	
PA (RIKA) INST PHYSICAL & CHEM RES.	

PI Hirabayashi Y, Ichikawa S;
 DR WPI: 97-036152/04.
 DR N-PSDB: T45357.
 PT New isolated human ceramide glucosyl:transferase - useful for prodn.
 CC of glucosyl:ceramide, a precursor of glycosphingolipid biosynthesis
 PS Claim 1, Page 13; 18pp; English.
 CC The present sequence represents a novel ceramide glucosyltransferase
 CC (gluc-1), which has catalytic activity for glucose transfer from
 CC UDP-glucose to ceramide. Ceramide glucosyltransferase catalyses the
 CC first glycosylation step of glycosphingoceramide to produce
 CC glucosylceramide, a precursor to more than 300 glycosphingolipid
 CC biosynthetases. The sequence can be used in biotechnology and
 CC biotechnology areas as laboratory agents, and also as preparatory
 CC agents for the manufacture of pharmacological active substances. It
 CC is expected to have use in therapeutic and preventive treatment of
 CC diseases based on reduced expression of ceramide
 CC glucosyltransferase.
 SQ Sequence 394 AA.

Query Match 10.1%; Score 63; DB 1; Length 394;
 Best Local Similarity 19.2%; Pred. No. 14;

Matches 25; Conservative 10; Mismatches 19; Indels 76; Gaps 7;

QY 6 LPAPVCDSDQNEC-----VGLGVAISAVVDYIRAAHEGYIEPEANGRLDALREWIT 59
 DB 284 LPATIIICEPISCEFAVSLIIGW-----AAHH-----VFRMDI 315
 QY 60 YNYTERSKRRDRRRSVCHARTWF-----CFPKYDY-----VRRS----- 96
 DB 316 MYFF-----MCHCLAMFLFDYIQLRGVGGTLCFSKLDYAVAMFIRESMIT 362
 QY 96 -----TWMDT 100
 DB 363 IFLSALMDPT 372

RESULT 3

ID Y05381 standard; Protein; 394 AA.

AC Y05381;
 DT 30-JUN-1999 (first entry)
 DE Human HCMV repressible gene protein, SEQ ID NO 35.
 KW HCMV repressible gene; crg; human; human cytomegalovirus; interferon;
 KW anti-viral therapy; anti-HCMV therapy; detection; diagnosis;
 OS Homo sapiens.
 KM drug screening.
 OS Homo sapiens.
 PN WO9913075-A2.
 PD 18-MAR-1999.
 PF 08-SEP-1998; U18638.
 PR 22-SEP-1997; US-059725.
 PR 08-SEP-1997; US-058180.
 PA (UYPR-) UNIV PRINCETON.
 PI Cong J, Schenk T, Zhu H;
 DR WPI: 99-243729/20.
 DR N-PSDB: X33956.
 PT New isolated human genes
 PS Claim 5; Page 174-176; 184pp; English.
 CC This sequence is encoded by a human gene of the invention, repressed in
 CC the presence of HCMV infection, designated HCMV-repressible genes (crg or
 CC crgs). The invention also relates to genes that are induced to express by
 CC both HCMV and interferon (IFN), designated HCMV-inducible genes (cig or
 CC cigs). The products can be used to obtain agents which can
 CC be used for anti-viral therapy, particularly anti-HCMV therapy. They can
 CC also be used for the development of drugs that would allow for higher
 CC dosage IFN treatments without the concomitant toxicity normally
 CC associated with administering high levels of IFN. The products can also
 CC be used for detection, diagnosis and drug screening.
 SQ Sequence 394 AA;

Query Match 10.1%; Score 63; DB 1; Length 394;
 Best Local Similarity 19.2%; Pred. No. 14;

Matches 25; Conservative 10; Mismatches 19; Indels 76; Gaps 7;

QY 6 LPAPVCDSDQNEC-----VGLGVAISAVVDYIRAAHEGYIEPEANGRLDALREWIT 59
 DB 284 LPATIIICEPISCEFAVSLIIGW-----AAHH-----VFRMDI 315
 QY 60 YNYTERSKRRDRRRSVCHARTWF-----CFPKYDY-----VRRS----- 96
 DB 316 MYFF-----MCHCLAMFLFDYIQLRGVGGTLCFSKLDYAVAMFIRESMIT 362
 QY 96 -----TWMDT 100
 DB 363 IFLSALMDPT 372

RESULT 4

ID W14748 standard; Protein; 2594 AA.

AC W14748;
 DT 13-MAY-1997 (first entry)
 DE IgG-Fc binding protein encoded by 7.8 kb fragment of pNV11-ST.
 KW Fragment 13; pNV11-ST; IgG-Fc binding protein; immunoglobulin; K17;
 KW human; colonic epithelium; monoclonal antibody; K9; probe.
 OS Homo sapiens.
 KM WO9527057-A1.
 PD 12-OCT-1995.
 PF 03-APR-1995; J00638.
 PR 01-APR-1994; JP-129487.
 PR 24-AUG-1994; JP-222547.
 PR 30-MAR-1995; JP-109927.
 PA (CHUS) CHUGAI SEIYAKU KK.
 PI Harada N, Morikawa M;
 DR WPI: 95-358632/46.
 DR N-PSDB: T63073.
 PT DNA derived from colonic epithelium encoding IgG-Fc binding protein
 PT - used in the mapping and analysis of IgG-Fc binding protein mRNA
 PS Claim 1; Page 71-84; 132pp; Japanese.
 CC This sequence is encoded by fragment 13 which is a NotI/KpnI fragment
 CC from pNV11-ST. This sequence represents a portion of the IgG-Fc binding
 CC protein of human colonic epithelium. This sequence was used in the
 CC isolation of the full length sequence given in W14748. mRNA isolated
 CC from human colonic epithelial tissue was used to prepare a cDNA library.
 CC This was screened using monoclonal antibodies K9 and K17 which bind to
 CC the large and small components of the binding protein. Active clones,
 CC see also T63077-81, were used to derive probes for screening a second
 CC DNA library from human colonic epithelial tissue.
 SQ Sequence 2594 AA;

Query Match 10.1%; Score 62.5; DB 1; Length 2594;
 Best Local Similarity 27.7%; Pred. No. 1.6e+02;
 Matches 23; Conservative 7; Mismatches 26; Indels 27; Gaps 3;

QY 5 ALPAPVCDSDQNECV-----GMLGVAISAVVDYIRAAHEGYIEPEANGRLDALREW 57
 DB 1551 ALSAPPC--ODGCAEGGCGDGFYXNGACVPIQCGCGYHNGYIEPEQVLLD----- 1604
 QY 58 IYNNYTERSKRRDRRRSVCH 80
 DB 1604 -----NCRQOCTCHA 1613

RESULT 5

ID W14749 standard; Protein; 5405 AA.

AC W14749;
 DT 13-MAY-1997 (first entry)
 DE IgG-Fc binding protein.
 KW Fragment 13; pNV11-ST; IgG-Fc binding protein; immunoglobulin; K17;
 KW human; colonic epithelium; monoclonal antibody; K9; probe.
 OS Homo sapiens.
 KM WO9527057-A1.
 PD 12-OCT-1995.

PF 03-APR-1995; J00638.
 PR 01-APR-1994; JP-129487.
 PR 24-AUG-1994; JP-222547.
 PR 30-MAR-1995; JP-109927.
 PA (CHUS) CHUGAI SEIRAKU KK.
 PI Harada N, Morikawa M;
 DR WPI; 95-358632/46.
 DR N-PSDB; T63074.
 PT DNA derived from colonic epithelium encoding IgG-Fc binding protein
 PT - used in the mapping and analysis of IgG-Fc binding protein mRNA
 PS Claim 3; Page 86-113; 132pp; Japanese.
 CC This sequence represents the IgG-Fc binding protein of human colonic
 CC epithelium. This sequence was isolated using the sequence given in
 CC W4748. mRNA isolated from human colonic epithelial tissue was used
 CC to prepare a cDNA library. This was screened using monoclonal antibodies
 CC K9 and K17 which bind to the large and small components of the binding
 CC protein. Active clones, see also T63077-81, were used to derive probes
 CC for screening a second DNA library from human colonic epithelial tissue.
 SQ Sequence 5405 AA;

Query Match 10.1%; Score 62.5; DB 1; Length 5405;
 Best Local Similarity 27.7%; Pred. No. 4e+02; Mismatches 26; Indels 27; Gaps 3;
 Matches 23; Conservative 7;

OY 5 ALPAPVCDSONECV-----GWLGVAYSANVDYIRAAHEGYIEPEARGRLDALREW 57
 DB 1551 ALSAPPC- QDGCAGCGCCSGFLYNGACVPIQCCGYNNGYEEQVRLD----- 1604
 OY 58 IYNYIYERSKRDRRRRSVCHA 80
 DB 1604 -----NCRQCCTCHA 1613

RESULT 6
 ID W97817
 AC W97817 standard; Protein; 150 AA.
 DT 21-MAY-1999 (first entry)
 DE Rat myelin oligodendrocyte glycoprotein.
 KW Myelin oligodendrocyte glycoprotein; MOG; rat; butyrophilin;
 KW multiple sclerosis; diagnosis; vaccine; transgenic animal; food;
 KW dairy product; milk.
 OS Rattus sp.
 PN W09905162-A1.
 PD 04-FEB-1999
 PR 22-JUL-1998; U15062.
 PR 22-JUL-1997; US-053376.
 PA (DYMA-) UNIV MARYLAND BALTIMORE.
 PI Linington C, Mather IH;
 DR WPI; 99-142845/12.
 PT New modified butyrophilin - not cross-reactive with myelin
 PT oligo-dendrocyte glycoprotein, useful in dairy products, vaccines,
 PT and assays for susceptibility to multiple sclerosis
 PS Disclosure: Page 46-47; 53pp; English.
 CC This polypeptide comprises the rat myelin oligodendrocyte
 CC glycoprotein (MOG). The invention relates to butyrophilin (BTN,
 CC see W97812-16), a major component of the milk fat globule membrane
 CC and member of the immunoglobulin (Ig) superfamily. The intermediate
 CC -type (Ig1) domain of BTN has a high degree of identity with the IgV
 CC domain of MOG. Immunisation with exogenous BTN induces the
 CC production of antibodies which cross-react with endogenous MOG, and
 CC this predisposes animals to pathological autoimmune responses such
 CC as multiple sclerosis (MS). The invention provides methods of
 CC modifying BTN to eliminate sequences that elicit this immune
 CC response to MOG, e.g. by removing or modifying the Ig1 domain of a
 CC non-human BTN (hnBTN) in a dairy product, or by producing a
 CC humanised chimeric BTN. The invention includes genetic manipulation
 CC of the endogenous BTN gene of an animal such that the human Ig1
 CC domain is introduced into the germline of the animal, and the
 CC animal expresses the humanised protein in its milk. Methods are
 CC provided for detecting susceptibility to MS in a human patient, from
 CC ingesting, and for identifying a human patient at risk of developing

CC a pathogenic autoimmune response to MOG on exposure to hnBTN is also
 CC possible, by screening the patient to identify polymorphisms in the
 CC gene with domains which are at least 50% identical to hnBTN Ig1
 CC domains (claimed).
 SQ Sequence 150 AA;

Query Match 10.0%; Score 62; DB 1; Length 150;
 Best Local Similarity 37.5%; Pred. No. 5.8;
 Matches 15; Conservative 5; Mismatches 20; Indels 0; Gaps 0;

OY 19 VGMGVAYSANVDYIRAAHEGYIEPEARGRLDALREWI 58
 DB 37 VGMVRSFSPRVHLYRNGKDDAQAPEYGRTELLKESI 76

RESULT 7
 ID W76041
 AC W76041 standard; Protein; 1661 AA.
 DT 20-NOV-1998 (first entry)
 DE Hydra head activator binding protein.
 KW Head activator binding protein; hydra; screening assay; agonist;
 KW antagonist; treatment; neurodegenerative disease.
 OS ChoroHydra viridissima.
 FH Key Location/Qualifiers
 FT Peptide 1..29
 FT /label= signal
 FT 30..1661
 FT /label= head_activating_binding_protein
 FT 80..84
 FT /note= "protease site"
 FT 85..99
 FT /label= N-terminus
 FT 121..737
 FT /label= VPS10
 FT /note= "As given in specification"
 FT 738..1002
 FT /label= 6x8
 FT /note= "As given in specification"
 FT 1003..1041
 FT /note= "EGF-like region as given in the specification"
 FT 1042..1367
 FT /label= 7x4
 FT /note= "As given in specification"
 FT 1368..1543
 FT /note= "Fibronectin type III"
 FT 1584..1606
 FT Domain /label= transmembrane

PN DE19808258-A1.
 PD 03-SEP-1998.
 PD 27-FEB-1998; 008258.
 PR 10-JUL-1997; DE-028490.
 PR 28-FEB-1997; DE-008134.
 PA (EVOT-) EVOTEC BIOSYSTEMS GMBH.
 PI Hampe W, Schaller C;
 DR WPI; 98-468562/41.
 DR N-PSDB; V46335.
 PT Hydra head activator binding protein - useful in screening assay for
 PT agonists or antagonists
 PS Claim 3; Page 6-11; 21pp; German.
 CC This sequence represents a head activator binding protein isolated from
 CC ChoroHydra viridissima by chromatographic methods. The protein is used
 CC in a screening assay for agonists or antagonists comprising determining
 CC the binding properties of candidate substances with the protein. The
 CC assay can be used to screen for substances useful for treating
 CC neurodegenerative diseases.
 SQ Sequence 1661 AA;

Query Match 9.9%; Score 61.5; DB 1; Length 1661;
 Best Local Similarity 31.3%; Pred. No. 1.2e+02;
 Matches 21; Conservative 4; Mismatches 25; Indels 17; Gaps 3;

```

OY      8 APPVCDSONECVWTL-----GVAYSAVVDYIRAAHEGYTIEPEARGRLDALR-----E 56
DB      499 SPSVYFSSNAGISMKRIFDGNNTYAYVD-----HGVATVGEKIGMTSYLKRSDGNT 552
OY      57 WYIYNXY 63
DB      553 WYSYNXY 559

RESULT  8
W10344  W10344 standard; Protein; 2763 AA.
AC      W10344;
DE      15-APR-1997 (first entry)
KW      MDW-B; viral resistance; disease resistance; transgenic plant;
KW      monococt; P3 proteinase; Nta proteinase; Nib replicase;
KW      RNA-dependent RNA polymerase; coat protein; Zea mays; sorghum;
KW      sugarcane; Saccharum officinale.
OS      Maize dwarf mosaic virus strain B.
FH      Key
FH      Location/Qualifiers
FT      1..377
FT      /label= HC-Pro
FT      /note= "portion of the helper component-P2
FT      proteinase"
FT      263..336
FT      /note= "MDW-B HC-Pro domain"
FT      378..791
FT      /label= P3-proteinase
FT      /note= "claimed polypeptide (Claim 20)"
FT      792..1430
FT      /label= Cylindrical inclusion protein
FT      /note= "claimed polypeptide (Claim 20)"
FT      880..1010
FT      /note= "conserved helicase domain"
FT      1431..1483
FT      /label= K2
FT      /note= "6 kDa protein"
FT      1484..1914
FT      /label= Nta-proteinase
FT      /note= "claimed polypeptide (Claim 20)"
FT      1915..2435
FT      /label= Nib-replicase
FT      /note= "claimed polypeptide (Claim 20)"
FT      2266..2268
FT      /note= "conserved motif characteristic of
FT      RNA-dependent replicases"
FT      2436..2763
FT      /label= Coat-protein
FT      144
FT      /note= "unidentified amino acid"
FT      704
FT      /note= "unidentified amino acid"
FT      712
FT      /note= "unidentified amino acid"
FT      829
FT      /note= "unidentified amino acid"
FT      834
FT      /note= "unidentified amino acid"
FT      843
FT      /note= "unidentified amino acid"
FT      847
FT      /note= "unidentified amino acid"
FT      852
FT      /note= "unidentified amino acid"
FT      1346
FT      /note= "unidentified amino acid"
FT      1362
FT      /note= "unidentified amino acid"
FT      2077
FT      /note= "unidentified amino acid"
FT      misc_difference 2385

```

```

FT      /note= "unidentified amino acid"
PN      MO9702352-A1.
PD      23-JAN-1997.
PF      20-JUN-1996; E02673.
PR      30-JUN-1995; US-496944.
PA      (CIBA ) CIBA GEIGY AG.
PI      Dietz JM, Law MD;
DR      WPI: 97-108965/10.
DR      N-PSDB: T47073.
FT      Chimeric gene for imparting viral resistance to plants - contains
FT      sequence modified to express non-translatable mRNA, or non-coat
FT      viral protein
PS      Claim 20; Page 31-44; 64pp; English.
CC      The sequence of the polypeptide encoded by the polycistronic mRNA
CC      (T47073) of maize dwarf mosaic virus strain B (MDW-B) is given in
CC      W10344. New chimeric genes comprise a monocotyledonous plant
CC      promoter linked to a modified nucleic acid sequence derived from
CC      the MDW-B genome. The modification is such that mRNA is translated
CC      to a truncated protein (pref. smaller than 200 amino acids), no
CC      translation of mRNA occurs or the transcribed mRNA lacks the
CC      translation initiation codon or includes a premature stop codon.
CC      Expression of the chimeric gene inhibits infection of plants (pref.
CC      sorghum, sugarcane, esp. maize) by MDW. The transgenic plants
CC      display an inheritable resistance trait.
SQ      Sequence 2763 AA;

Query Match      9.9%; Score 61.5; DB 1; Length 2763;
Best Local Similarity 29.8%; Pred. No. 2.3e+02;
Matches 14; Conservative 9; Mismatches 17; Indels 7; Gaps 1;

OY      34 RAAHEGYTIEPEARGRLDALREMYTYNTERSKRRDRRRSYCHA 80
DB      2313 RGKTEGITYPKLEKRIVALTEW-----DRSNLPBRLDAICAA 2352

RESULT  9
R22421  R22421 standard; Protein; 118 AA.
AC      R22421;
DE      29-JUL-1992 (first entry)
KW      Antibody produced from HuVH1conHMFg1.
KW      Anti-HMFg specificity; complementarity determining region; CDR;
KW      human milk fat globule; polymorphic epithelial mucin; PEM.
OS      Homo sapiens.
FH      Key
FH      Location/Qualifiers
FT      31..35
FT      /label= CDR1
FT      /note= "from mouse heavy chain"
FT      50..65
FT      /label= CDR2
FT      /note= "from mouse heavy chain"
FT      95..103
FT      /label= CDR3
FT      /note= "from mouse heavy chain"
FT      region
FT      MO9204380-A.
FT      19-MAR-1992.
FT      05-SEP-1991; G01511.
FT      07-SEP-1990; GB-019553.
FT      (UNIL ) UNILEVER PLC.
FT      Verhoeven M;
FT      WPI: 92-114305/14.
FT      N-PSDB: 023349.
DR      Synthetic specific binding agent and reshaped human antibody -
FT      specific for human polymorphic epithelial mucin for treatment and
FT      in-vivo diagnosis of PEM-producing cancers
PS      Claim 11; Fig 12; 62pp; English.
CC      The reshaped antibody sequence was deduced from the sequence of a
CC      reshaped human variable heavy chain gene which had a human class I
CC      framework but contained mouse CDRs. To reshape a human heavy chain,
CC      three fragments, each one contg. a murine CDR, were assembled to
CC      form gene HuVH1conHMFg1. The reshaped gene was inserted into a
CC      pSV vector to produce a plasmid for transformation of E. coli.

```


CC The antibodies produced are essentially human so are unlikely
 CC to cause adverse reactions. The reshaped molecule can be used
 CC to treat PEM-producing cancers (e.g. of the breast, ovary, uterus
 CC or lung) and in in-vivo diagnosis.
 CC See also R22419-22.
 SQ Sequence 118 AA:

Query Match 9.7%; Score 60; DB 1; Length 118;
 Best Local Similarity 23.2%; Pred. No. 7.4;
 Matches 23; Conservative 15; Mismatches 39; Indels 22; Gaps 4;

QY 23 GVAYSANVDVIRAAHHEGV-----YIEPARGRLALREMIYNYTTERS 67
 DB 26 GTFSAYWIEWRQAGKLEWVGEILPGSNNSRYNE-KFKGRVYTRDTSTNTAYMELS 84
 QY 68 KRDRRRRSVCARWFCFRKYDYVRSIWHDTTNTIS 106
 DB 85 SLRSE-----DTAVYYCARSYDFAMFAWVGQGLTVTS 117

RESULT 10
 RS4758
 ID RS4758; standard; Protein; 118 AA.
 AC RS4758;
 DT 06-DEC-1994 (first entry)
 DE Humanised HMG1 heavy chain variable region.
 KW Single chain Fv fragment; Scfv; Gene Therapy; Monoclonal Antibody;
 OS Homo sapiens.

FT misc_difference 5 Location/Qualifiers
 FT misc_difference 11.12 /label- humanised by replacing Gln with Val
 FT misc_difference 20 /label- humanised by replacing Leu and Met with
 FT misc_difference 25 /label- humanised by replacing Ile with Val
 FT misc_difference 38 /label- humanised by replacing Thr with Ser
 FT misc_difference 40 /label- humanised by replacing Lys with Arg
 FT misc_difference 43 /label- humanised by replacing Arg with Ala
 FT misc_difference 48 /label- humanised by replacing His with Lys
 FT misc_difference 67.68 /label- humanised by replacing Ile with Val
 FT misc_difference /label- humanised by replacing Lys and Ala
 FT misc_difference 70 with Arg and Val
 FT misc_difference 72 /label- humanised by replacing Phe with Val
 FT misc_difference 76 /label- humanised by replacing Ala with Arg
 FT misc_difference 82 /label- humanised by replacing Ser with Thr
 FT misc_difference 87 /label- humanised by replacing Gln with Glu
 FT misc_difference 91 /label- humanised by replacing Thr with Arg
 FT misc_difference 97 /label- humanised by replacing Ser with Thr
 FT misc_difference 113 /label- humanised by replacing Ser with Ala
 FT misc_difference 118 /label- humanised by replacing Pro with Leu
 FT misc_difference 118 /label- humanised by replacing Ala with Ser
 FT protein 1.118 /label- humanised HMG1 monoclonal antibody
 FT note- "heavy chain variable region"
 FT region 1.30

FT region /label- FR 1
 FT region 31.35 /label- CDR 1
 FT region 36.49 /label- FR 2
 FT region 50.66 /label- CDR 2
 FT region 67.98 /label- FR 3
 FT region 99.107 /label- CDR 3
 FT region 108.118 /label- FR 4
 FT region /label- FR 4
 FT WO9410323-A.
 PN 11-MAY-1994.
 PD 04-NOV-1993: G02267.
 PR 04-NOV-1992: GB-023084.
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Epenetos AA, Spooner RA;
 DR N-PSDB; 064819.
 DR WPI; 94-167477/20.
 PT Virus with modified binding moiety specific for the target cells
 PT - used to deliver genes for gene therapy and cancer treatment
 PS Example 4; Page 74; 110pp; English.
 CC RS4758 shows an humanised HMG1 heavy chain variable region. The
 CC mouse HMG1 is changed at specific residues within the framework region.
 CC It is used to generate a scfv with the humanised HMG1 light chain
 CC variable region (R63804).
 CC The invention is concerned with the fusion of an scfv with the Ad5
 CC fibre such that the fusion sequence can be used to target cells for
 CC gene therapy.
 SQ Sequence 118 AA:

Query Match 9.7%; Score 60; DB 1; Length 118;
 Best Local Similarity 23.2%; Pred. No. 7.4;
 Matches 23; Conservative 15; Mismatches 39; Indels 22; Gaps 4;

QY 23 GVAYSANVDVIRAAHHEGV-----YIEPARGRLALREMIYNYTTERS 67
 DB 26 GTFSAYWIEWRQAGKLEWVGEILPGSNNSRYNE-KFKGRVYTRDTSTNTAYMELS 84
 QY 68 KRDRRRRSVCARWFCFRKYDYVRSIWHDTTNTIS 106
 DB 85 SLRSE-----DTAVYYCARSYDFAMFAWVGQGLTVTS 117

RESULT 11
 Y13128
 ID Y13128 standard; Protein; 133 AA.
 AC Y13128;
 DT 22-JUN-1999 (first entry)
 DE Human secreted protein encoded by 5' EST SEQ ID NO: 142.
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
 KW forensic; gene therapy; chromosome mapping; signal peptide;
 KW upstream regulatory sequence; cytokine activity; cell proliferation;
 KW differentiation; haematopoiesis regulation; tissue growth regulation;
 KW reproductive hormone regulation; chemotactic; chemokine; haemostatic;
 KW thrombolytic; anti-inflammatory; tumour inhibition.
 OS Homo sapiens.
 PN WO906552-42.
 PD 11-FEB-1999.
 PE 31-JUL-1998; IB1236.
 PR 01-AUG-1997; US-9052223.
 PA (GEST) GENSET.
 PI Ducleier A, Dumas M,Line Edwards J, Lacroix B;
 DR WPI; 99-153782/13.
 DR N-PSDB; X51928.
 PT New isolated brain-derived nucleic acids - used to develop products
 PT which may have cytokine, immune, regulatory, haematopoiesis
 PT regulating, anti-inflammatory or tumour inhibition activity
 PS Claim 34; Page 522-523; 577pp; English.
 CC X51767 to X52019 represent 5' expressed sequence tags (ESTs) for human

secreted proteins, and encode the proteins given in Y12987 to Y13219, respectively. The proteins given represent the signal peptide and an N-terminal fragment of a secreted protein. The nucleic acid sequences can be used for producing secreted human gene products. They can also be used to develop products for diagnosis and therapy. The proteins obtained may have cytokine activity, cell proliferation/differentiation activity, haematopoiesis regulating activity, tissue growth regulating activity, reproductive hormone regulating activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, receptor/ligand activity, anti-inflammatory activity, tumour inhibition activity or other activities. The products can be used in forensic, gene therapy and chromosome mapping procedures. The sequences can also be used for obtaining corresponding promoter sequences. The nucleic acids encoding the signal peptide can be used for directing extracellular secretion of a polypeptide or the insertion of a polypeptide into a membrane, or importing a polypeptide into a cell.

Sequence 133 AA;

Query Match 9.7%; Score 60; DB 1; Length 133;
Best Local Similarity 41.7%; Pred. No. 8.6;
Matches 10; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 60 YNYTERKRRDRRRRCYCHARTW 83
DB 32 YAYHERKERDPERHALCDLGPW 55

RESULT 12

W56774 standard; Protein; 323 AA.

AC W56774; 13-OCT-1998 (first entry)

DE Homo sapiens PSP1 partial sequence.

KW PS-1; presenilin; presenilin-1; PSP-1; Alzheimer's disease;

OS serine protease; neurodegeneration; predisposition; diagnosis.

PN EP-828003-A2.

PD 11-MAR-1998.

PF 26-AUG-1997; 306501.

PR 13-DEC-1996; US-032875.

PR 06-SEP-1996; US-025436.

PR 25-OCT-1996; US-027873.

PA (SMIR) SMITHKLINE BEECHAM CORP.

PI Browne MJ, Clinkenbeard HE, Creasy CL, Karan EH,

PI Liyi GP, Southan CD;

DR WPI; 98-161101/15.

DR N-PSDB; V29522.

PT Nucleic acids encoding human serum protease protein(s) - used for

diagnosing pre-disposition to Alzheimer's disease, etc.

PS Example 2; Page 18-19; 65pp; English.

CC The sequence is that of a fragment of the serine

protease PSP1. This can be used to identify modulators

of serine protease activity and also to diagnose a

condition associated with lack of one of the serine proteases

or a genetic predisposition to neurodegeneration in a patient,

preferably predisposition to Alzheimer's disease.

Sequence 323 AA;

Query Match 9.7%; Score 60; DB 1; Length 323;
Best Local Similarity 32.6%; Pred. No. 25;
Matches 14; Conservative 8; Mismatches 13; Indels 8; Gaps 1;

QY 2 VLPALPAPVCDQNECVGMLGVAVSAYVDYVIRAAHEGYIE 44
DB 77 VLAAYVSPPPASPRSQ-----YNFADYVEKTPAPVAYIE 111

RESULT 13
W56774 standard; Protein; 377 AA.

AC W56774;
DT 13-OCT-1998 (first entry)
DE Homo sapiens PSP1-3 consensus sequence.
KW PS-1; presenilin; presenilin-1; PSP-1; Alzheimer's disease;
KW serine protease; neurodegeneration; predisposition; diagnosis.
OS Homo sapiens.
PN EP-828003-A2.
PD 11-MAR-1998.
PF 26-AUG-1997; 306501.
PR 13-DEC-1996; US-032875.
PR 06-SEP-1996; US-025436.
PR 25-OCT-1996; US-027873.
PA (SMIR) SMITHKLINE BEECHAM CORP.
PI Browne MJ, Clinkenbeard HE, Creasy CL, Karan EH,
PI Liyi GP, Southan CD;

DR WPI; 98-161101/15.

DR N-PSDB; V29537.

PT Nucleic acids encoding human serum protease protein(s) - used for

diagnosing pre-disposition to Alzheimer's disease, etc.

PS Claim 7; Page 41-42; 65pp; English.

CC The sequence is that of the consensus sequence of PSP1-3

serine protease which can be used to identify modulators of

serine protease activity and also to diagnose a condition

associated with lack of one of the serine proteases or

a genetic predisposition to neurodegeneration in a patient,

preferably predisposition to Alzheimer's disease.

Sequence 377 AA;

Query Match 9.7%; Score 60; DB 1; Length 377;
Best Local Similarity 32.6%; Pred. No. 31;
Matches 14; Conservative 8; Mismatches 13; Indels 8; Gaps 1;

QY 2 VLPALPAPVCDQNECVGMLGVAVSAYVDYVIRAAHEGYIE 44
DB 131 VLAAYVSPPPASPRSQ-----YNFADYVEKTPAPVAYIE 165

RESULT 14

ID R13506 standard; Protein; 413 AA.

AC R13506;

DT 25-OCT-1991 (first entry)

DE P.denitrificans COB L.

KW cob gene; corrinoid; descobalocorrinoid; cor gene.

OS Pseudomonas denitrificans.

PN WO911518-A.

PD 08-AUG-1991.

PF 30-JAN-1991; F00054.

PR 31-JAN-1990; FR-001137.

PA (RHON) RHONE-POULENC BIOCH.

PI Blanche F, Meron B, Crouzet J, Debussche L, Levy-Schli S;

PI Thibaut D;

DR WPI; 91-252650/34.

DR N-PSDB; Q13285.

PT New Polypeptide(s) involved in cobalamin and cobamide

biosynthesis - and DNA encoding them, for amplification of

PT cobalamin, esp. coenzyme B12 prodn.

PS Claim 17; Fig 16; 289pp; French.

CC This sequence corresponds to one of 24 polypeptides obtained from

P.denitrificans and implicated in the biosynthesis of cobalamines

and/or cobamides, specifically in catalysing the transfer of a

methyl group to positions C1, C5, C11, C15 or C17 in the conversion

of precorrin-3 to cobyrinic acid a,c-diamide. It is encoded by part

of the 8.7kb EcoRI-EcoRI fragment of plasmid pX1367. The plasmid was

isolated from a P.denitrificans genomic DNA bank constructed in

vector pX159.

CC See Q13284-Q13288.

CC Sequence 413 AA;

Query Match 9.7%; Score 60; DB 1; Length 413;

Best Local Similarity 48.6%; Pred. No. 34;
Matches 17; Conservative 2; Mismatches 16; Indels 0; Gaps 0;

QY 22 LGVAYSADVIRAAHEGYIEPEARGRLDALRE 56
DB 205 LGLVHPLNVCALIEVADEGARILPLLAGRDALFE 239

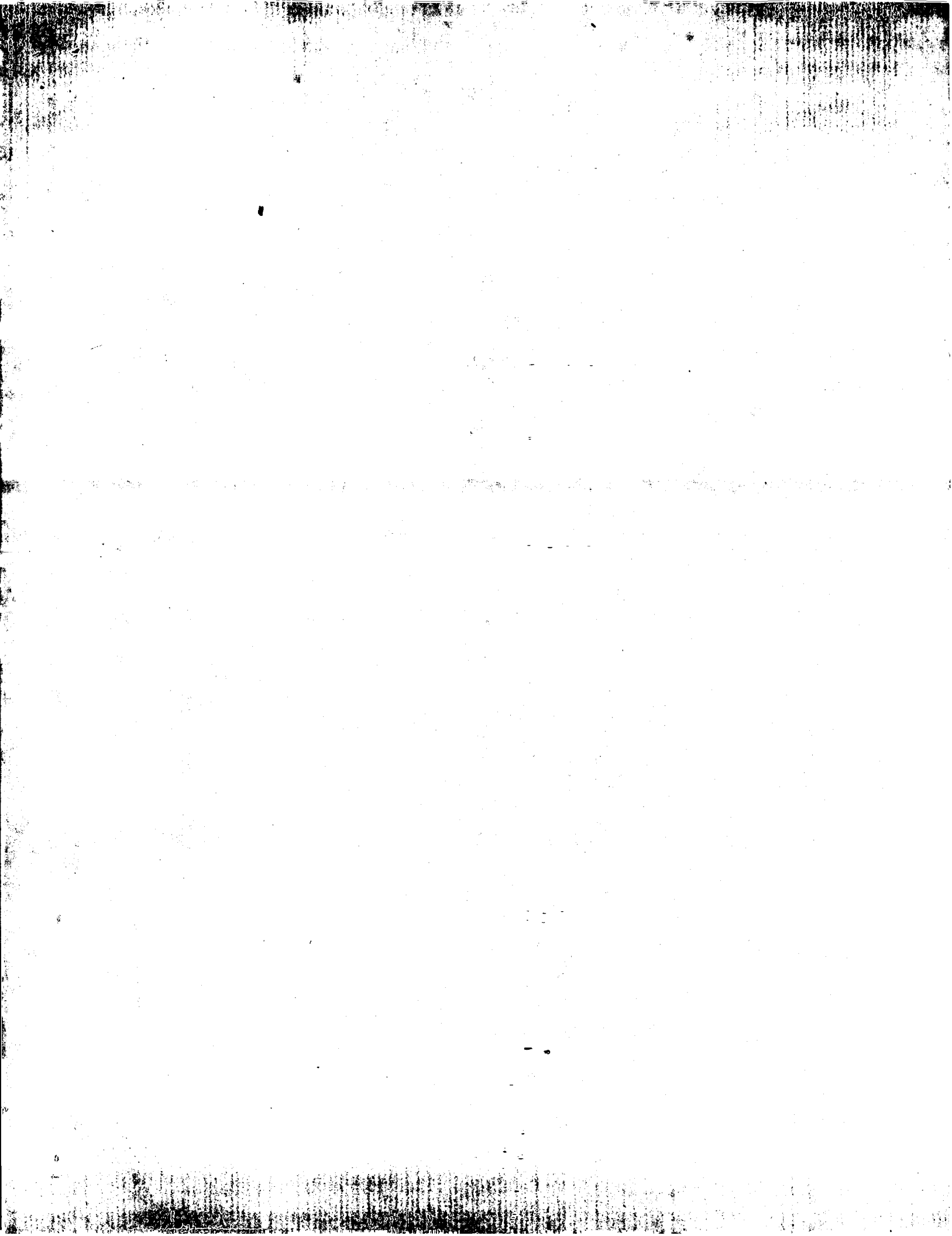
RESULT 15

W56776
ID W56776 strand; Protein; 436 AA.
AC W56776;
DT 13-OCT-1998 (first entry)
DE Homo sapiens PSP1-4 consensus sequence.
KW PS-1; presenilin; presenilin-1; PSP1-4; Alzheimer's disease;
KW serine protease; neurodegeneration; predisposition; diagnosis.
OS Homo sapiens.
PN EP-828003-A2.
PD 11-MAR-1998.
PF 26-APR-1997; 306501.
PR 13-DEC-1996; US-032875.
PR 06-SEP-1996; US-025436.
PR 25-OCT-1996; US-027873.
PA (SMIR) SMITHKLINE BEECHAM CORP.
PI Browne MJ, Clinkenbeard HE, Creasy CL, Kattan EH,
PI Livl GP, Southan CD;
DR WPI; 98-161101/15.
DR N-PDB; V29538.
PT Nucleic acids encoding human serum protease protein(s) - used for
PT diagnosing pre-disposition to Alzheimer's disease, etc.
PS Claim 7; Page 44-45; 65pp; English.
CC The sequence is that of the consensus sequence of PSP1-4
CC serine protease which can be used to identify modulators of
CC serine protease activity and also to diagnose a condition
CC associated with lack of one of the serine proteases or
CC a genetic predisposition to neurodegeneration in a patient,
CC preferably predisposition to Alzheimer's disease.
SQ Sequence 436 AA;

Query Match 9.7%; Score 60; DB 1; Length 436;
Best Local Similarity 32.6%; Pred. No. 37;
Matches 14; Conservative 8; Mismatches 13; Indels 8; Gaps 1;

QY 2 VLPALPAPVPCDSQNECVGLGVAYSADVIRAAHEGYIE 44
DB 131 VLAAPVSPSPSPASPSQ-----YNFADVVEKTAAPAVYIE 165

Search completed: June 23, 2000, 10:03:38
Job time: 3366 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: June 23, 2000, 10:04:33 ; Search time 45.41 Seconds

(without alignments)
36.245 Million cell updates/sec

Title: US-09-214-478-4

Perfect score: 621
Sequence: 1 MYLPALPAPPCDSQNECVG.....SIMHDTTNTISVSASHSYQ 114Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 145308 seqs, 14437401 residues

Total number of hits satisfying chosen parameters: 145308

Minimum DB seq length: 0
Maximum DB seq length: 1000000Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

1: Issued Patents, AA:*
2: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	63	10.1	394	1	US-08-663-713A-2
2	61.5	9.9	2763	3	US-08-496-944-2
3	60	9.7	118	2	US-08-428-257A-78
4	59	9.5	649	3	US-08-995-441B-109
5	58.5	9.4	890	2	US-08-483-101-14
6	57	9.2	135	2	US-08-860-174A-4
7	57	9.2	215	2	US-08-659-251-8
8	57	9.2	215	4	PCT-US96-11445-8
9	57	9.2	274	2	US-08-860-174A-12
10	57	9.2	1019	1	US-08-296-014A-4
11	57	9.2	1019	2	US-08-595-405-4
12	57	9.2	1019	2	US-08-877-620-4
13	57	9.2	15281	2	US-08-471-119A-2
14	56.5	9.1	115	2	US-08-379-057-31
15	56.5	9.1	115	2	US-08-428-197-42
16	56.5	9.1	115	4	PCT-US93-10555-42
17	56.5	9.1	788	2	US-08-918-914-4
18	56	9.0	622	2	US-08-458-146-2
19	56	9.0	622	2	US-08-459-065-2
20	56	9.0	1015	1	US-08-537-210A-1
21	56	9.0	2471	1	US-08-185-432-16
22	56	9.0	2471	1	US-08-083-590A-19
23	55.5	8.9	137	1	US-08-331-398A-61
24	55.5	8.9	137	2	US-08-331-387B-61
25	55.5	8.9	137	2	US-08-753-804A-60
26	55	8.9	247	4	PCT-US94-10257A-2
27	55	8.9	489	4	PCT-US94-10521-9
28	55	8.9	503	2	US-08-394-189B-2
29	55	8.9	503	4	PCT-US93-05701-19

ALIGNMENTS

30	55	8.9	503	4	PCT-US93-05705-2	Sequence 2, Appl1
31	54.5	8.8	487	1	US-08-249-112-4	Sequence 4, Appl1
32	54.5	8.8	487	1	PCT-US95-06556-4	Sequence 4, Appl1
33	54.5	8.8	611	1	US-08-386-727-4	Sequence 4, Appl1
34	54.5	8.8	611	2	US-08-600-452A-4	Sequence 4, Appl1
35	54	8.7	138	2	US-08-379-057-14	Sequence 14, Appl1
36	54	8.7	144	2	US-08-116-778E-36	Sequence 36, Appl1
37	54	8.7	144	2	US-08-438-562-36	Sequence 36, Appl1
38	54	8.7	144	2	US-08-483-528B-100	Sequence 100, App
39	54	8.7	379	2	US-08-871-074-26	Sequence 26, Appl1
40	54	8.7	489	1	US-08-489-733-6	Sequence 6, Appl1
41	54	8.7	489	2	US-08-993-581B-6	Sequence 6, Appl1
42	54	8.7	491	1	US-08-489-733-5	Sequence 5, Appl1
43	54	8.7	491	2	US-08-993-581B-5	Sequence 5, Appl1
44	54	8.7	760	1	US-08-230-491A-2	Sequence 2, Appl1
45	54	8.7	760	1	US-08-519-280A-2	Sequence 2, Appl1

RESULT 1

US-08-663-713A-2
Sequence 2, Application US/08663713A

Patent No. 5780284

GENERAL INFORMATION:

APPLICANT: YOSHIO HIRABAYASHI et al.

TITLE OF INVENTION: CERAMIDE GLUCOSYLTRANSFERASE

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wenderoth, Lind & Ponack

STREET: 805 Fifteenth Street, N.W., #700

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 Inch, 500 kb

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/663,713A

FILING DATE: June 14, 1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Warren M. Cheek, Jr.

REGISTRATION NUMBER: 33,367

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-371-8850

TELEFAX:

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 394 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-663-713A-2

Query Match 10.1% Score 63; DB 1; Length 394;

Best Local Similarity 19.2% Pred. No. 6.8; Indels 76; Gaps 7;

Matches 25; Conservative 10; Mismatches 19; Indels 76; Gaps 7;

QY 6 LPAPPCDSQNECVG-----VGMGLGAVSAVVIVRAAHGEGVTEPEARGRLDLREWIY 59

DB 284 LPATPCIPICPCFASLIIQM-----AAHH-----VFRRDI 315

OY 60 YNTERSKRRDRRRSVCHARTWF-----CFKRYD-----VRS----- 96
DB 316 MYFE-----MCHCLAMFIDYIQLGVOGGLCFKSLDPAVAMFIRESENTIY 362
OY 96 -----IWHDT 100
DB 363 IFLSALMDPT 372

RESULT 2

US-08-496-944-2
Sequence 2, Application US/08496944
Patent No. 6040496

GENERAL INFORMATION:

APPLICANT: Law, Marcus D
APPLICANT: Dietz, Jon M
TITLE OF INVENTION: Use of Translationally altered RNA to
TITLE OF INVENTION: Confer Resistance to Maitze Dwarf Mosaic Virus and Other
TITLE OF INVENTION: Monocytiledonous Plant Viruses
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA

ZIP: 10532
COUNTRY: USA

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/496,944
FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: CGC 1814

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 2763 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-496-944-2

Query Match 9.7%; Score 61.5; DB 3; Length 2763;
Best Local Similarity 29.8%; Pred. No. 1.3e+02;
Matches 14; Conservative 9; Mismatches 17; Indels 7; Gaps 1;

OY 34 RAAAHGVYIEPARGRLALREWIYNYTERSKRRDRRRSVCH 80
DB 2313 RGIKYGIIIPKIKERIVAILDEM-----DSSNIPERHLEATCA 2352

RESULT 3

US-08-428-257A-78
Sequence 78, Application US/08428257A
Patent No. 585808

GENERAL INFORMATION:

APPLICANT: Spooner, Robert A.
APPLICANT: Epenetos, A.A.

TITLE OF INVENTION: Compounds to target cells
NUMBER OF SEQUENCES: 80

CORRESPONDENCE ADDRESS:

ADDRESSEE: Jules E. Goldberg
STREET: 261 Madison Avenue
CITY: New York
STATE: NY
COUNTRY: USA

ZIP: 10016-2391
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,257A
FILING DATE: 07/05/95
CLASSIFICATION: 514
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 118 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-428-257A-78

Query Match 9.7%; Score 60; DB 2; Length 118;
Best Local Similarity 23.2%; Pred. No. 3.3;
Matches 23; Conservative 15; Mismatches 39; Indels 22; Gaps 4;

OY 23 GVAISAV-VDVIRAAAHGVY-----YIEPARGLDALREWIYNYTERS 67
DB 26 GYFSAYWIEWVROAFKGLWVGEILPGSSNNSRYNE-KFKGRVYVTRDTSTNTAYMELS 84

OY 68 KRDRRRRSVCHARTWFCFKRYDYVRSIWHDTNTIS 106
DB 85 SLRSE-----DTAVYYCARSTDFANFAIWGGCTLVYS 117

RESULT 4

US-08-996-441B-109
Sequence 109, Application US/08996441B
Patent No. 6023013

GENERAL INFORMATION:

APPLICANT: English, Leigh H.
APPLICANT: Brussock, Susan M.

APPLICANT: Malvar, Thomas M.
APPLICANT: Bryson, James W.

APPLICANT: Kulesza, Caroline A.
APPLICANT: Walters, Frederick S.

APPLICANT: Slatin, Stephen L.
APPLICANT: Von Terssch, Michael A.

APPLICANT: Romano, Charles
TITLE OF INVENTION: INSECT-RESISTANT TRANSGENIC PLANTS
NUMBER OF SEQUENCES: 113

CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433
CITY: Houston

STATE: Texas
COUNTRY: USA

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/996,441B

FILING DATE: 18-DEC-1997
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:
NAME: Mitchell, Barbara S.

REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: MECO:151

TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000

TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 109:

SEQUENCE CHARACTERISTICS:

LENGTH: 649 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-996-441B-109

Query Match 9.5%; Score 59; DB 3; Length 649;
Best Local Similarity 22.0%; Pred. No. 41;
Matches 22; Conservative 12; Mismatches 32; Indels 34; Gaps 4;

QY 16 NECVGLGVAVSANDVIRAAHGVYIEPEARGLALREMIYNTYTERSKRRRR 75
DB 239 NHAKW---YKGLDKRSTYE-----EWYKFNRY-----RREM 270

QY 76 SVCHARTFCFRKYD-----VRSIMHDTTNTISVS 109
DB 271 TLTVLDLITFPLVDVRYTKGVTELTDRVLDPIYAVN 310

RESULT 5

US-08-483-101-14
Sequence 14, Application US/08483101
Patent No. 5932715

GENERAL INFORMATION:

APPLICANT: Scott, June R.
APPLICANT: Caron, Judy

TITLE OF INVENTION: CS2 Proteins and Coding Sequences
NUMBER OF SEQUENCES: 17

CORRESPONDENCE ADDRESS:

ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201

CITY: Boulder
STATE: Colorado

COUNTRY: US
ZIP: 80303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,101

FILING DATE: 07-JUN-1995
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: Farber, Donna M.

REGISTRATION NUMBER: 33878
REFERENCE/DOCKET NUMBER: 6-95

TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080

TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:
LENGTH: 890 amino acids

TYPE: amino acid
STRANDEDNESS:

TOPOLOGY: unknown
MOLECULE TYPE: protein

HYPOTHETICAL: NO
US-08-483-101-14

Query Match 9.4%; Score 58.5; DB 2; Length 890;
Best Local Similarity 31.0%; Pred. No. 71;
Matches 18; Conservative 10; Mismatches 21; Indels 9; Gaps 3;

QY 19 VGMGLV--AVSANDV-----IRAAHGVYIEPEARGLALREMIYNTYTERSKR 69
DB 488 MGVGCGESYASLSIPKGNLSLAYSNT--STSVRYDAVSEYVYVYKGNKTR 543

RESULT 6

US-08-860-174A-4
Sequence 4, Application US/08860174A
Patent No. 5989830

GENERAL INFORMATION:

APPLICANT: DAVIS, Paul James
APPLICANT: VAN DER LOGT, Cornelis Paul Erik

APPLICANT: VERHOEIJEN, Martine Elisa
APPLICANT: WILSON, Steve

TITLE OF INVENTION: A BIFUNCTIONAL OR BIVALENT
TITLE OF INVENTION: ANTIBODY FRAGMENT ANALOGUE

NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:

ADDRESSEE: PILLSBURY, MADISON & SUTRO LLP
STREET: 1100 New York Avenue, N.W.

STREET: 9th Floor, East Tower
CITY: WASHINGTON, D.C.

STATE:
COUNTRY: UNITED STATES

ZIP: 20005-3918

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS DOS Text
SOFTWARE: MS WORD, WINDOWS 95, Ver. 4.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,174A

FILING DATE: June 16, 1997
PRIOR APPLICATION DATA: EP 95307332.7

FILING DATE: October 16, 1995
APPLICATION NUMBER: PCT/EP/96/03605

FILING DATE: August 14, 1996
INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 135 amino acids

TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-860-174A-4

Query Match 9.2%; Score 57; DB 2; Length 135;
Best Local Similarity 25.0%; Pred. No. 9.2;
Matches 18; Conservative 10; Mismatches 18; Indels 26; Gaps 4;

QY 56 EMIIYNY-----YTERSKRR---DRRRSVCH-----ARTWCFRKYDYVR 94
DB 68 EMIGITIPYNDGTRKNEKFKATLTSKSSSTAYMELSLTSEDSAVYCSRRFY--- 125

QY 95 SIMHDTTNTIS 106
DB 125 --WGQGTTVVS 134

RESULT 7

US-08-659-251-8
Sequence 8, Application US/08659251
Patent No. 5883081

GENERAL INFORMATION:

APPLICANT: Kraus, Guenter
APPLICANT: Wong-Staal, Flossie

APPLICANT: Talbot, Randy
APPLICANT: Poeschla, Eric

TITLE OF INVENTION: Isolation of No. 5883081el HIV-2 Proviruses
NUMBER OF SEQUENCES: 50

CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

GENERAL INFORMATION:
APPLICANT: The Regents of the University of California
TITLE OF INVENTION: Isolation of Novel HIV-2 Proviruses
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: Robbins, Berliner & Carson
STREET: 201 N. Figueroa Street, 5th Floor
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90012-2628
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/11445
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Berliner, Robert
REGISTRATION NUMBER: 20,121
REFERENCE/DOCKET NUMBER: 3535-399C1

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Floppy disk
3  COMPUTER: IBM PC compatible
4  OPERATING SYSTEM: PC-DOS/MS-DOS
5  SOFTWARE: MS WORD, WINDOWS 95, VER. 4.0
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: US/08/860, 174A
8  FILING DATE: June 16, 1997
9  PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: EP 95307332.7
11 FILING DATE: October 16, 1995
12 APPLICATION NUMBER: PCT/EP/96/03605
13 FILING DATE: August 14, 1996
14 INFORMATION FOR SEQ ID NO: 12:
15     SEQUENCE CHARACTERISTICS:
16         LENGTH: 274 amino acids
17         TYPE: amino acid
18         TOPOLOGY: linear
19     MOLECULE TYPE: protein
20     OS-08-860-174A-12

```

Query Match 9.28; Score 57; DB 2; Length 274;

Best Local Similarity 25.0%; Pred. No. 23;
Matches 18; Conservative 10; Mismatches 18; Indels 26; Gaps 4;

QY 56 EWYIYV-----YERSRR-----DRRRSICH-----ARTWCFKRYVVR 94
Db 207 EWGTYIPYNDGTYKFKATLTSDKSSYAMELSLTSEDNAVYICSRFDY--- 264

QY 95 SIWDTTNTIS 106
Db 264 --WGQTYTVS 273

RESULT 10
US-08-296-014A-4
Sequence 4, Application US/08296014A
Patent No. 5716834

GENERAL INFORMATION:

APPLICANT: Ding, Jeak Ling
TITLE OF INVENTION: The Cloned Factor C cDNA of the
TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinoscopus
TITLE OF INVENTION: roundicauda and Purification of Factor C Proenzyme
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22042

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296,014A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1781-105P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1019 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-296-014A-4

Query Match 9.2%; Score 57; DB 1; Length 1019;
Best Local Similarity 23.2%; Pred. No. 1.3e+02;
Matches 23; Conservative 18; Mismatches 40; Indels 18; Gaps 5;

QY 21 WLGVAYSADVIRAAHEGYIEPEARGLDALREWIYNY-TERSKRRDRRRSVCH 79
Db 362 WGTATYHELSSVCRAIHAQKL--PNSGAVHYVNGPYSDFLGSLNGIKSELSKL-- 418

QY 80 ARTWCFKRYDYVRSI-----WHDTTNTISYS 109
Db 418 ARSF---RFDYVRSYTAGKSGCPDGWFEVDENCYVTS 452

RESULT 11
US-08-596-405-4
Sequence 4, Application US/08596405
Patent No. 5858706

GENERAL INFORMATION:

APPLICANT: Ding, Jeak Ling
TITLE OF INVENTION: The Cloned Factor C cDNA of the
TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinoscopus
TITLE OF INVENTION: roundicauda and Purification of Factor C Proenzyme
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22042

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/596,405
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1781-105P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
TELEX: 248345

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 1019 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-596-405-4

Query Match 9.2%; Score 57; DB 2; Length 1019;
Best Local Similarity 23.2%; Pred. No. 1.3e+02;
Matches 23; Conservative 18; Mismatches 40; Indels 18; Gaps 5;

QY 21 WLGVAYSADVIRAAHEGYIEPEARGLDALREWIYNY-TERSKRRDRRRSVCH 79
Db 362 WGTATYHELSSVCRAIHAQKL--PNSGAVHYVNGPYSDFLGSLNGIKSELSKL-- 418

QY 80 ARTWCFKRYDYVRSI-----WHDTTNTISYS 109
Db 418 ARSF---RFDYVRSYTAGKSGCPDGWFEVDENCYVTS 452

RESULT 12
US-08-877-620-4
Sequence 4, Application US/08877620
Patent No. 5985590

GENERAL INFORMATION:

APPLICANT: Ding, Jeak Ling
TITLE OF INVENTION: The Cloned Factor C cDNA of the
TITLE OF INVENTION: Singapore Horseshoe Crab, Carcinoscopus
TITLE OF INVENTION: roundicauda and Purification of Factor C Proenzyme
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 8110 Gatehouse Road, Suite 500 East
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22042

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/877,620
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/596,405
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1781-105P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050

TELEX: 248345
INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 1019 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-877-620-4

Query Match 9.2%; Score 57; DB 2; Length 1019;
Best Local Similarity 23.2%; Pred. No. 1.3e+02;
Matches 23; Conservative 18; Mismatches 40; Indels 18; Gaps 5;

QY 21 WGVASAVVDVIRAAHEGYIEPANGRLALREMIYNTY-TERSKRRDRRRSYCH 79
DB 362 WGTATYHEHSYCRALHAGKL--PNSGAVHYVNNNGPYSDPLGSDLNIGKSELKSL-- 418
QY 80 ARTWCFRRKYDVRSI-----WHDYTNNTISYVS 109
DB 418 ARSF-----RFDYVRSSTAGKSGCPDGFVENDVCYVYS 452

RESULT 13
US-08-471-119A-2
Sequence 2, Application US/08471119A
Patent No. 5827706
GENERAL INFORMATION:
APPLICANT: Leitner, Ernst
APPLICANT: Schneider, Elisabeth
APPLICANT: Schoergerdorfer, Kurt
APPLICANT: Weber, Gerhard
TITLE OF INVENTION: Cyclosporin Synthetase
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5827706artis Corporation
STREET: 59 Route 10
CITY: East Hanover
STATE: New Jersey
COUNTRY: USA
ZIP: 07936
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,119A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kassenoft, Melvyn
REGISTRATION NUMBER: 26,389
REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 503 8474

TELEFAX: 201 503 8807
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 15281 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
HYPOTHETICAL: NO
AMT-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Tolypocladium niveum
STRAIN: ATCC 34921
US-08-471-119A-2

Query Match 9.2%; Score 57; DB 2; Length 15281;
Best Local Similarity 30.5%; Pred. No. 4.5e+03;
Matches 18; Conservative 7; Mismatches 22; Indels 12; Gaps 2;

QY 11 VCDSE-----ONECVGWLGVASAVVDVIRAAHEGYIEPEA--RGRLDLAKRM 57
DB 3844 ICESYDLDGDAKSDNDYVSWLSARSNVAVASISALDVLDAQEAQFVEISCAROW 3902

RESULT 14
US-08-379-057-31
Sequence 31, Application US/08379057
Patent No. 5876950

GENERAL INFORMATION:
APPLICANT: Sladak, Anthony W.
APPLICANT: Hollenbaugh, Diane L.
APPLICANT: Gilliland, Lisa K.
APPLICANT: Gordon, Marcia L.
APPLICANT: Bajorath, Jürgen
APPLICANT: Aruffo, Alejandro A.
TITLE OF INVENTION: Monoclonal Antibodies Specific For
TITLE OF INVENTION: Different Epitopes of Human gp99 and Methods For Their Use
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bristol-Myers Squibb Company
STREET: 3005 First Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98121

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/379,057
FILING DATE: 26-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Poor, Brian W.
REGISTRATION NUMBER: 32,928
REFERENCE/DOCKET NUMBER: ON0133-
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 727-3670
TELEFAX: (206) 727-3601
INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:
LENGTH: 115 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-379-057-31

Mon Jun 26 09:10:58 2000

us-09-214-478-4.rai

Page 8

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 23, 2000, 09:17:09 ; Search time 592.38 Seconds
(without alignments)
2360.583 Million cell updates/sec

Title: US-09-214-478-3

Sequence: 345
1 atggtcttcacgcctcttc.....cggcgactcgcgtacagtag 345

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 segs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : EST:
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*

45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*
54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*
59: gb_est33:*
60: gb_est34:*
61: gb_est35:*
62: gb_est36:*
63: gb_est37:*
64: gb_est38:*
65: em_est27:*
66: em_est28:*
67: em_est29:*
68: em_est30:*
69: gb_est39:*
70: gb_est40:*
71: gb_est41:*
72: gb_est42:*
73: gb_est43:*
74: gb_est44:*
75: em_est31:*
76: em_est32:*
77: em_est33:*
78: em_est34:*
79: gb_est45:*
80: gb_est46:*
81: gb_est47:*
82: gb_gsa1:*
83: gb_gsa2:*
84: gb_gsa3:*
85: gb_gsa4:*
86: em_gsa1:*
87: em_gsa2:*
88: em_gsa3:*
89: em_gsa4:*
90: gb_gsa5:*
91: gb_gsa6:*
92: gb_gsa7:*
93: gb_gsa8:*
94: gb_gsa9:*
95: em_gsa5:*
96: em_gsa6:*
97: em_gsa7:*
98: em_gsa8:*
99: em_gsa9:*
100: em_gsa10:*
101: em_gsa11:*
102: gb_gsa10:*
103: gb_gsa11:*
104: em_gsa12:*
105: gb_gsa12:*
106: gb_gsa13:*
107: gb_gsa14:*
108: gb_gsa15:*
109: gb_gsa16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result

Query

SUMMARIES

No.	Score	Match Length	DB	ID	Description
1	303	87.8	630	79	AM276836 xp6661.x
2	138.4	40.1	426	79	AM265066 xp60503.x
3	35.4	10.3	303	26	W18889 mc04g02.r1
4	35.4	10.3	345	36	AA656067 vs49h11.x
5	35.4	10.3	348	35	AA562033 x125609.r1
6	35.4	10.3	389	34	AA509726 vs20f06.r1
7	35.4	10.3	453	26	W30475 mc18a09.r1
8	35.4	10.3	477	27	AA041778 xp041778.mj02g10.r1
9	35.4	10.3	479	27	W82078 mc96g07.r1
10	35.4	10.3	504	39	AA874208 vx03004.r1
11	35.4	10.3	533	60	AI1875404 vj41e10.x
12	35.4	10.3	543	29	AA140104 mc93a05.r1
13	35.4	10.3	547	31	AA290263 vc90g08.r1
14	35.4	10.3	551	60	AA290263 vc90g08.r1
15	35.4	10.3	556	43	AI195511 vj22a03.x
16	35.4	10.3	555	30	AA195511 vj50g10.y
17	35.4	10.3	655	47	AA217816 mm98g09.r1
18	35.4	10.3	676	42	AI481623 v24d411.x
19	35.4	10.3	945	26	AA115246 vj43a10.y
20	35	10.1	684	44	AI322378 m06d10.r1
21	34.8	10.1	581	102	AO362256 mgxh00012
22	34.8	10.1	656	105	AO325183 mgxh00022
23	34.8	10.1	803	91	AO891719 HS_3100.BS
24	34	9.9	697	69	AM1071116 um19a01.r1
25	33.8	9.8	312	43	AI195796 vj50g10.x
26	33.8	9.8	339	81	AA1412524 uq42a01.x
27	33.8	9.8	489	33	AA404076 vq78f05.r1
28	33.8	9.8	662	28	AA123016 mp82e05.r1
29	33.6	9.7	530	63	AI394038 701456727
30	33.4	9.7	909	82	CMS00507
31	32.8	9.5	661	109	AO638467 927p1-12
32	32.2	9.3	483	27	W82481 m04d01.r1
33	32.2	9.3	545	45	AI389550 m2d9b06.y
34	31.4	9.1	241	32	AA377776 EST90394
35	31.4	9.1	748	82	AG001595 Hmo sap1
36	31.2	9.0	449	63	AM006018 w291p08.x
37	31.2	9.0	659	72	AM163374 au9d410.y
38	31.2	9.0	710	92	AO944247 Sheared D
39	31	9.0	256	25	W06958 za91a01.r1
40	31	9.0	360	91	AO841382 r136655b
41	31	9.0	407	91	AO841382 r136655b
42	30.8	8.9	465	22	AI4844 ym25a04.r1
43	30.8	8.9	739	82	AG001559 Hmo sap1
44	30.6	8.9	354	64	AM039817 EST282290
45	30.6	8.9	752	37	AA703929 ag46c01.s

ALIGNMENTS

RESULT 1
AM276836 LOCUS
DEFINITION xp66911.x1 NCI-CGAP_Ov39 Homo sapiens cDNA clone IMAGE:2745380 3' similar to SW:5413_ADE02 P03240 PROBABLE EARLY E4 I3 KD PROTEIN.
mRNA sequence.
ACCESSION AM276836
VERSION AM276836
KEYWORDS AM276836.1 GI:6663866
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo. "
REFERENCE 1 (bases 1 to 630)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Nov 22, 1999 this sequence version replaced gi:6462296.
CONTACT: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550

Email: Robert.Strausberg@nih.gov
Tissue Procurement: Andrew Bercchuck M.D., John Gillespie M.D.,
Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I M A G E Consortium/LLNL at:
www-bio.llnl.gov/dbip/image/image.html

Possible reversed clone: similarity on wrong strand
Possible reversed clone: polyT not found
Seq primer: -400p from Gibco
High quality sequence stop: 422.

FEATURES
Source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2745380"
/clone_1lb="NCI-CGAP_Ov39"
/sex="Female"
/tissue_type="papillary serous ovarian metastasis"
/lab_host="DH10B"
/notes="Organ: ovary; Vector: PAMp10; CDNA made by oligo-dT priming. Non-directionally cloned into the upg sites of PAMp10. Size-selected on agarose gel, average insert size 500 bp. Primary library, non-amplified. CDNA Library Preparation: David B. Kitzman, Ph.D (NCI). Reference: Kitzman et al. (1996) Cancer Research 56:5380-5383."

```

BASE COUNT	154 a	134 c	173 g	168 t	1 others
ORIGIN					

Query Match	87.8%;	Score 303;	DB 79;	Length 630;
Best Local Similarity	88.1%;	Prod No. 1	50-95;	

Best local similarity 50.18; Freq.NO. 1.35e-03;
Matches 306; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Y 1 atggttcttcagctcttccgcctcctccgtgtgtgactcgcaagaatgtgtagt 60

b 319 ATGGTCTCCAGCTCTCCCGCTCCTCCGCTGTGTACTCGCAGACGATGTGAGGT 370

61 tgcctggtgtgacctatctctgcggtgtgtgatgttatcaaggcagcgcgcataaagga 120

b 379 TGGCTGGGTGGCTTATTCCTGGGTGGTGATGTTATCANGGCAGCGCGCATGAGGA 430

121 gttacatagaaacccgaagccaggggcccctgtagtgcctttagagaagtgcatactac 180

b 439 GTTACATAGACCCGGAAGCCAGGGGGCCCTGGATGCTTGGAGAGCTGGATATACTAC 499

181 aactactacacagagcgatcttaagcgcgcgagacccggaacgcagatctgttgcacgcc 24

b 499 AACTACTACACAGCGAGCTAAGCGACGAGACCGAGACCGAGATCTGTTGTACGCC 550

241 cgcacctggttttgcttcaggaataatgactacgtccggcgcttccatttgcatgaact 300

b 559 CGACCTGGTTTGGCTTCAGAAATATGACTACGTGGGGCTTGCATTGGCATGACACT 618

Y 301 acgaccaacacg 312

b 619 ACGACCAACAGC 630

ESMUT 2

AW265066
mCCTG
AW265065
425 bp
mRNA
EST
28-DEC-1999

DEFINITION	NCBI	CDNA	IMAGE
similar to SW:EA11 ADE02 P03241 PROBABLE EARLY E4 11 KD PROTEIN	qx60g03.x1	Homo sapiens	2755066

ACCESSION [1] ;, mRNA sequence.
AW265066

AM265066.1 GI:6641882
EST.

human.

ORGANISM	Homo sapiens
EMBRYO	Embryos: Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 426) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) On Jul 7, 1999 this sequence version replaced gi:5406301. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov unknown library type Possible reversed clone: similarity on wrong strand Seg primer: -40bp from Gbpco High quality sequence stop: 425. Location/Qualifiers 1..426
FEATURES	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2755060" /clone_lib="NCI CGAP Co22" /tissue_type="colonic adenocarcinoma" /lab_host="DH10B" /note="Organ: colon; Vector: PAMP10; cDNA made by oligo-dT priming. Non-directionally cloned into the UDG sites of PAMP10. Size-selected on agarose gel, average insert size 500 bp. Primary library: non-amplified. cDNA library Preparation: David B. Kitzman, Ph.D (NCI). Reference: Kitzman et al. (1996) Cancer Research 56:5380-5383."
BASE COUNT	103 a 80 c 124 t 1 others
ORIGIN	
Query Match	40 1%; Score 138.4; DB 79; Length 426;
Best Local Similarity	98.6%; Pred. No. 1.7e-33; Indels 0; Gaps 0;
Matches 139; Conservative	0; Mismatches 2;
OY	1 atggtttcttcgcgcgtcttcgcgcgtccgcgcgtgtgtactgcgacgaacgaatgtgtagt 60
Db	286 atggtttcttcgcgcgtcttcgcgcgtccgcgcgtgtgtactgcgacgaacgaatgtgtagt 345
OY	61 tggcgtgtgtgtgtgtatcttcgcgcgtgtgtgtgtatcatcagcagcgcgcgtatgaaga 120
Db	346 tggcgtgtgtgtgtgtatcttcgcgcgtgtgtgtgtatcatcagcagcgcgcgtatgaaga 405
OY	121 gtttaccatagaacccgaagcc 141
Db	406 gtttaccatagaacccgaagcc 426
RESULT 3	
LOCUS	W18889 303 bp mRNA EST 10-SEP-1996
DEFINITION	mc04g02.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone IMAGE:337970 5' similar to SW:BI4_RAT P5287 RAS-RELATED PROTEIN RAB-14. [1] ; mRNA sequence.
ACCESSION	W18889
VERSION	W18889.1 GI:1294599
KEYWORDS	EST.
SOURCE	house mouse. Mus musculus
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 303) Marrs,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Gelsel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,R., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE	The WashU-HMIT Mouse EST Project

JOURNAL
COMMENT

Unpublished (1996)
On May 9, 1995 this sequence version replaced gI:802677.

Contact: Marra Mouse EST Project
WashU-HMNI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL ; contact the IMAGE Consortium (linfoimage.lnl.gov) for further information.
MG1:219370

FEATURES
SOURCE

location/Qualifiers
 1..303
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone_id="IMAGE:337970"
 /_clone_1id="Soares mouse p3NMF19.5"
 /dev_stage="19.5 dpc total fetus"
 /lab_host="DHIOB (ampicillin resistant)"
 /note="Vector: pRTT3D (Pharmacia) with a modified polylinker site_1. Not I, Site_2. Eco RI; 1st strand cDNA was primed with a Not I - oligo(dt) primer [5', TGTATTCACATCTGAAGTGCGGAGCCGCCCATTTTTTTTTT 3'] , adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRTT3 vector . (Pharmacica). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA was kindly provided by Dr. Minoru Ko (Wayne State University)." 5' t

BASE COUNT
ORIGIN

106 a 61 c 77 g 59 t

Query Match 10.3%; Score 35.4; DB 26; Length 303;
Best Local Similarity 52.3%; Pred. No. 0.61;
Matches 78; Conservative 0; Mismatches 71; Indels 0; Gaps 0.

Qy 75 ttattctgcggcgtgagtgattacaggcgccgatgaaggagttaacataagaacc 134
 | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8 TCAAACTGCAGATTTCGGCAATACAGCAGCGCACAGACC GTTACAGCGCTTACACGAGCT 67

Qy 135 cgaaagccaggsggcgccttgcgttccttgtagagagtggatatctaactactaacagaca 194
 | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 68 ACTATTAGAGAGACCTGCAGGTGCCCTCATGGCTTAATGACATCCAGAMAGAATACTATATA 127

Qy 195 gcgacctaaagcggcgagaccggagacgccca 223
 | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 128 ACCA CTTAAGCACAGCTGGTTGACAGACGCA 156

RESULT 4
LOCUS AA656067
DEFINITION vs49hl1_r1 Stratigene mouse Toell 93731 Mus musculus CDNA clone
 IMAE:1149669 S' similar to SW-RL4_RAT P35287 RAS-RELATED PROTEIN
ACCESSION RAB-14.; mRNA sequence.
VERSION AA656067
KEYWORDS AA656067.1 GI:2592221
SOURCE EST.
ORGANISM house mouse.
 Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Schirognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubugue,T,
 Geisler,S., Kuubaba,T., Lacey,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

TITLE
JOURNAL
COMMENT

Unpublished (1996)
The WashU-HM1 Mouse EST Project
On Sep 12, 1996 this sequence version replaced g1:1393029.
Contact: Mairra M/Mouse EST Project
WashU-HM1 Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu

JRES
source

FEATURES

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through ILLN; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI: A247220
Possible reversed clone: similarity on wrong strand
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 449.

Source

BASE COUNT	163 a	85 c	120 g	111 t
ORIGIN				

Query Match	10.38;	Score 35.4;	DB 27;	Length 479;
Best Local Similarity	52.38;	Pred. No. 0.74;		
Matches 78;	Conservative	0;	Mismatches 71;	Indels 0;
			Gaps	0

OY	75	ttaattccgcggttgtagatgatcctacggcgagcgccabtgaaagtgaatttcataaac	13
Db	88	TCAACCTGCAGATTGGGATACAGAGAGGCCAGAGCCGTTCAAGCGGTTCACCGACT	147
OY	135	cgaagccagaaggggcgccctcgatgctttggagaagatgatatactacaactacacaga	194
Db	148	ACRTATAGGAAGACTCAGTGGCTCATGGTGTTGATCACATCCAGAAAGTMCATATA	207
OY	195	ggcatctaagcggcagaaaccggaagca	223
Db	208	ACCACATTAAAGCAGCTGGTTGACAGACCCA	236

ORGANISMS

RESULT	10
AA874208	
LOCUS	AA874208 504 bp mRNA EST 19-MAR-1998
DEFINITION	vx30304.r1 Soares-thymus_2nbmt Mus musculus cDNA clone
IMAGS:	126314.5' similar to SW:R14_RAT P35287 RAS-RELATED PROTEIN
RAB-14.1' mRNA sequence.	
ACCESSION	AA874208
VERSION	AA874208.1 GI:2978897
KEYWORDS	EST.
SOURCE	house mouse.

RESULT 14
 A1784945 551 bp mRNA EST 02-JUL-1999
 LOCUS u122a03.x1 Sugano mouse kidney m1a Mus musculus cDNA clone
 DEFINITION IMAGE:1920652.3' similar to SW:RB14_RAT P35287 RAS-RELATED PROTEIN
 RAB-14., mRNA sequence.
 ACCESSION A1784945
 KEYWORDS A1784945.1 GI:5332696
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 551)
 Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
 Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
 Person,B., Swaller,T., Gibbons,M., Pape,D., Harrey,N., Schurk,R.,
 Rutter,E., Kohn,S., Shin,T., Jackson,T., Cardenas,M., McCann,R.,
 Waterston,R. and Wilson,R.
 The WashU-NCI Mouse EST Project 1999
 Unpublished (1999)
 On Jun 22, 1998 this sequence version replaced g1:3247058.
 CONTACT: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.wustl.edu
 This clone is available royalty-free through LML; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:976644
 Possible reversed clone: similarity on wrong strand
 Seq primer: custom primer used
 High quality sequence stop: 481.

FEATURES

source
 1..551
 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /clone="IMAGE:1920652"
 /clone_1lb="Sugano mouse kidney m1a"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: kidney; Vector: pME18s-FL3; Site.1: DraIII
 (CACCTGTG); Site.2: DraIII (CACCACTG); 1st strand cDNA
 was primed with an oligo(dT) primer
 (ATGTCGCTTTTCTTTTCTTTT); double-stranded cDNA was
 ligated to a DraIII adaptor (TTGTGGCTACTG), digested
 and cloned into distinct DraIII sites of the pME18s-FL3
 vector (5' site CACCTGTG, 3' site CACCACTG). XhoI should
 be used to isolate the cDNA insert. Size selection was
 performed to exclude fragments <1.5kb. Library
 constructed by Dr. Sumio Sugano (University of Tokyo
 Institute of Medical Science). Custom primers for
 sequencing: 5' end primer CTCTGCTCTTAAAGCTGCG and 3' end
 primer CGACCTGCGCTCGAGCAACA."
 BASE COUNT 150 a 123 c 92 g 186 t
 ORIGIN

Query Match 10.3%; Score 35.4; DB 60; Length 551;
 Best Local Similarity 52.3%; Pred. No. 0.79;
 Matches 78; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
 Oy 75 ttattctgctgtgtgtatttatcagcgagcgatgaagagttacatagaacc 134
 Db 401 TCAAACTGCGAGATTGGATACAGACGAGCGATTCAGACCGCTTACAGAGCT 342
 Oy 135 cgaagcagcgagcgagcctgtgattcttgagagagtgatatactacacactacacaga 194
 Db 341 ACTATGAGAGAGCTGCGAGTGCCTCATGTGTGTATGCATTCACCAAGAGATACATATA 282

Oy 195 ggcattcaagcgagcgagaccgagacga 223
 Db 281 ACCACTTAAGACAGCTGTTGACAGACGCA 253
 RESULT 15
 A1195511 556 bp mRNA EST 14-OCT-1998
 LOCUS u150g10.y1 Sugano mouse liver m1a Mus musculus cDNA clone
 DEFINITION IMAGE:1885890.5' similar to SW:RB14_RAT P35287 RAS-RELATED PROTEIN
 RAB-14., mRNA sequence.
 ACCESSION A1195511
 KEYWORDS A1195511.1 GI:3748117
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 556)
 Marra,M., Hillier,L., Allen,M., Bowers,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)
 On Jan 19, 1998 this sequence version replaced g1:2286940.
 CONTACT: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.wustl.edu
 This clone is available royalty-free through LML; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:970214
 Possible reversed clone: similarity on wrong strand
 Seq primer: custom primer used
 High quality sequence stop: 509.

FEATURES

source
 1..556
 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /clone="IMAGE:1885890"
 /clone_1lb="Sugano mouse liver m1a"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: liver; Vector: pME18s-FL3; Site.1: DraIII
 (CACCTGTG); Site.2: DraIII (CACCACTG); 1st strand cDNA
 was primed with an oligo(dT) primer
 (ATGTCGCTTTTCTTTTCTTTT); double-stranded cDNA was
 ligated to a DraIII adaptor (TTGTGGCTACTG), digested
 and cloned into distinct DraIII sites of the pME18s-FL3
 vector (5' site CACCTGTG, 3' site CACCACTG). XhoI should
 be used to isolate the cDNA insert. Size selection was
 performed to exclude fragments <1.5kb. Library
 constructed by Dr. Sumio Sugano (University of Tokyo
 Institute of Medical Science). Custom primers for
 sequencing: 5' end primer CTCTGCTCTTAAAGCTGCG and 3' end
 primer CGACCTGCGCTCGAGCAACA."
 BASE COUNT 138 a 161 c 139 g 118 t
 ORIGIN

Query Match 10.3%; Score 35.4; DB 43; Length 556;
 Best Local Similarity 52.3%; Pred. No. 0.8;
 Matches 78; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
 Oy 75 ttattctgctgtgtgtatttatcagcgagcgagatgaagagttacatagaacc 134

Mon Jun 26 09:10:54 2000

us-09-214-478-3.rst

Page 10

Db 408 TCAACCTCAGATTGGGATTCACACAGGGCGAGAGCGGTTACAGCCGATTACACGGAGCT 467

Oy 135 cgaagccagagggggcgccctcgatgtttgagagagtgagatatactacacactactacacaga 184

Db 468 ACTATAGAGGAGCGTCAAGTCGGCTCATGTGCTATGACATGCCAGAAAGTATGATATA 527

Oy 195 ggcacatcaagcggcgagacacggagacaga 223

Db 528 AOCACCTTAACGACGCTGGTTGACACAGCGA 556

Search completed: June 23, 2000, 09:17:14
Job time: 12882 sec

QY	61	tgctcgggtgtgctgctattcttcgcgcgttggtgtgagtgtctatcaaggcgcgcgcctatgaaga	120
Db	1533	tgcctgggtgtgctgctattcttcgcgcgttggtgtgagtgtctatcaaggcgcgcgcctatgaaga	1474
QY	121	gtttacatagaaccccggaagccagggggcgccctgtagtgttttgaagagttgatatctac	180
Db	1473	gtttacatagaaccccggaagccagggggcgccctgtagtgttttgaagagttgatatctac	1414
QY	181	aactctcacagagcagttcttaagcggcgagagaccggaaacagcagatctgtttgcagcc	240
Db	1413	aactctcacagagcagttcttaagcggcgagagaccggaaacagcagatctgtttgcagcc	1354
QY	241	cgcacctggtttgttcttaaggaatatgactacgctcgcgcgttccatttgcatgacact	300
Db	1353	cgcacctggtttgttcttaaggaatatgactacgctcgcgcgttccatttgcatgacact	1294
QY	301	acgacacaacagatctcggtgtgtctcgcgcgacatccgtacagtag	345
Db	1293	acgacacaacagatctcggtgtgtctcgcgcgacatccgtacagtag	1249
RESULT 2			
ID	V32370/C		
AC	V32370 standard; DNA: 8710 BP.		
DT	23-SEP-1998 (first entry)		
DE	Complete sequence of the pE4/Hygro plasmid.		
KE	Circular; adenovirus type 5; pE4/Hygro plasmid; structural protein;		
KW	Complementation: E4 regulatory protein, gene therapy; HIV; tumour;		
KW	Huntington's disease; Tay-Sachs disease; sickle cell disease;		
KW	early gene; ds.		
OS	Synthetic.		
FS	Key		
FT	Location/Qualifiers		
FT	complement (707..3820)		
FT	/*tag= a		
FT	/note= "E4 regulatory gene"		
FT	3830..6470		
FT	/*tag= b		
FT	/note= "Hygromycin resistance gene"		
PN	W09813499-A2.		
PD	02-APR-1998.		
PF	24-SEP-1997; E05251.		
PR	25-SEP-1986; US-719806.		
PA	(NOVS) NOVARTIS AG.		
PA	(SCRI) SCRIPS RES INST.		
PI	Memmow GR, Von Seggern DJ;		
DR	WPI; 98-230709/20.		
PT	Adenoviral vectors - which lack DNA encoding for structural protein		
PT	or fibre protein used particularly for gene therapy		
PS	Example 1; Pages 63-74; 170pp; English.		
CC	The present sequence is that of a pE4/Hygro plasmid used in the		
CC	method of the invention. The plasmid contains an adenovirus type 5		
CC	E4 regulatory gene and a hygromycin resistant gene. The invention		
CC	provides adenoviral vectors having deletions of all or part of		
CC	various gene sequences encoding adenoviral structural proteins.		
CC	Deletions in the structural proteins would allow a reduced risk of		
CC	wild-type virus contamination and would also allow packaging of foreign		
CC	DNA in such vectors for a variety of diagnostic and therapeutic		
CC	applications. The adenoviral vectors having deletions in the structural		
CC	gene regions are produced by cellular complementation of these		
CC	adenoviral genes. Therefore, the pE4/Hygro plasmid was used as a		
CC	complementation plasmid which was introduced into a host cell line		
CC	where parts of the E4 gene region would be stably inserted into the		
CC	host cell chromosomes. The resulting E4 deficient plasmid can be used		
CC	as a gene delivery vector. The vectors can be used for diagnosis or		
CC	gene therapy, e.g. for treating conditions characterised by		
CC	hyper-proliferative cells (e.g. tumours), genetic diseases (e.g.		
CC	Huntington's disease, Tay-Sachs disease), or sickle cell disease), or		
CC	infections (e.g. HIV infection). They can also be used for in vitro		
CC	production of biologically active proteins. 2141 G; 1941 T;		
CC	Sequence 8710 BP; 2263 A; 2365 C; 2141 G; 1941 T;		

Query Match	Similarity	Score	DB	Length
Best Local	Similarity	100.0%	Pred. No. 8.2e-105	Length 8710.
Matches	345	Conservative	0	Mismatches
			0	Indels
			0	Gaps
QY	1	atggtcttcacagctctccgcctccctccgtgtgtaactcagagaacgatgtgaagt	60	
Db	2382	ATGGTCTTCCAGCTCTTCCGCTCCCTCCGCTGTGTGACACCGAGAACGATGTGAAGT	2323	
QY	61	tggcttggtgtgctctatcttcggtgtgtgatatgtatacaggcagcgcgcatgaagga	120	
Db	2322	TGGCTGGGTGTGGCTTATTCCTGCGGTGTGTGAGATTATCAGGGCAGCGCATGAGGA	2263	
QY	121	gtttcacaagaaccggaagccgagggtggcgctgtgattgtttgaaagtgatatactac	180	
Db	2262	GTTTTACATAGAACCCGAGGCCAGGGGGGCGCTGTGATGCTTTGAGAGTGGATATAC	2203	
QY	181	aactactacagagcgcatctaaacggtcgagacggagagacgcagatcttgttcacgcc	240	
Db	2202	AACACTACTACAGAGCGATCTTAAGCGGCGAGACCGGAGAGCGAGATCTTTGTACGCC	2143	
QY	241	cgcacgtggtttgtgtcttcagaagaaatatgactacgttcggcgcttccatttggcatgac	300	
Db	2142	CGCACGTGGTTTCTTCACGAAATATGATGATCGTCCGGGCTTCATTGGCATGACACT	2083	
QY	301	acgaccacaacgcatcgtggtgtgttcggcgacatccgtacagtag	345	
Db	2082	ACGACCACACGATCTCGGTTGTCTCGGCGCATCCGTACAGTAG	2038	
RESULT	3			
V32375/c				
ID	V32375	standard; DNA; 10610 BP.		
AC	V32375			
DT	29-SEP-1998	(first entry).		
DE	Complete sequence of the pE4/Fiber plasmid.			
KW	Circular; adenovirus type 5; pE4/Fiber plasmid; structural protein;			
KM	complementation; fiber protein; gene therapy; HIV; tumour; AD5;			
KN	early gene; Huntington's disease; Tay-Sachs disease;			
KM	sickle cell disease; E4 regulatory gene; ds.			
OS	Synthetic.			
FS	Key	Location/Qualifiers		
FT	misc_feature	complement (21..3149)		
FT		/*tag- a		
FT	sig_peptide	/note= "AD5 E4 regulatory gene"		
FT		4051..4366		
FT		/*tag- b		
FT	misc_feature	/note= "AD5 leader sequence"		
FT		4372..6124		
FT		/*tag- c		
FT		/note= "AD5 fiber gene"		
PN	W09813499-A2.			
PD	02-APR-1998.			
PF	24-SEP-1997.	E05251.		
PR	25-SEP-1996.	US-719806.		
PA	(NOVS) NOVARTIS AG.			
PA	(SCRI) SCRIPPS RES INST.			
PI	Memerow GR, Von Seggern DJ;			
DR	WPI: 98-230709/20.			
PT	Adenoviral vectors - which lack DNA encoding for structural protein			
PS	Example 1; Pages 131-145; 110pp; English.			
CC	The present sequence is that of a pE4/Fiber plasmid used in the			
CC	method of the invention. The plasmid contains an adenovirus type 5 (AD5)			
CC	fiber gene controlled by a CMV promoter, an AD5 E4 gene and an adenovirus			
CC	type 2 (AD2) tripartite leader sequence upstream of the fiber gene. The			
CC	invention provides adenoviral vectors having deletions of all or part of			
CC	various gene sequences encoding adenoviral structural proteins and/or			
CC	early region proteins. Deletions in these proteins would allow a			
CC	reduced risk of wild type virus contamination and would also allow			
CC	packaging of foreign DNA in such vectors for a variety of diagnostic and			
CC	therapeutic applications. The adenoviral vectors having deletions in			
CC	the structural and/or early gene regions are produced by cellular			
CC	...			


```

FT      /*tag= C
FT      /note= "region deleted in plasmid pdeprp and
FT      PBHG1delP (Claim 51)."
FT      misc_feature 11067.12513
FT      /tag= d
FT      /note= "region deleted in plasmid
FT      PAYBdelPdelPVPARNA+L13 and
FT      PBHG1delPdelPVPARNA+L13 (Claim 53)."
FT      WO9817783-A1.
FT      30-APR-1998.
FT      23-OCT-1997; U19541.
FT      23-OCT-1996; US-735609.
FT      23-OCT-1996; US-735609.
FT      (UNM1) UNIV MICHIGAN.
FT      Amalfitano A, Chamberlain JS, Hartigan-O'Connor DJ,
FT      Hauser MA, Kumar-Singh R;
FT      WPI: 98-261485/23.
FT      New adenoviral recombinant plasmid(s) - comprise sequences provided
FT      for expression of large foreign DNA fragments, used for, e.g. gene
FT      therapy of genetic disease(s)
FT      PS Claim 45; Page 87-102; 139pp; English.
FT      CC This nucleotide sequence comprises plasmid PBHG1 that consists of
FT      an E1-deleted adenovirus (Ad) genome. It contains a deletion of
FT      Ad5 (see also V07228) from bp 188-1339 (0.5-3.7 m.u.). This
FT      deletion removes the packaging signals as well as E1 sequences.
FT      PBHG1 also contains a large deletion within the E3 region (bp
FT      27865-30995, 77.5-86.2 m.u.). Nucleotide 8773 of PBHG1 is
FT      equivalent to nucleotide 7269 of Ad5. The large E3 deletion
FT      provides a large cloning capacity to Ad vectors based on PBHG1.
FT      CC Recombinant plasmids pdeprp and PBHG1delP (deleted for E1 and
FT      CC polymerase functions) and PAYBdelPdelPVPARNA+L13 and
FT      CC PBHG1delPdelPVPARNA+L13PBHG1 (deleted for E1, polymerase as
FT      CC preterminal protein functions), are specifically claimed. The
FT      CC invention provides improved adenoviral vectors and packaging
FT      CC within the E2b region of the adenoviral genome (see also V07261).
FT      CC These E2b-deleted virus are used in conjunction with novel cell
FT      CC lines that constitutively express E2b gene products. The invention
FT      CC also provides vectors deleted for all viral coding regions. These
FT      CC "gutted" vectors permit the transfer of large genes (e.g. up to
FT      CC 35 kb) to cells, as demonstrated by the transfer of the dystrophin
FT      CC gene to the muscle of mice. The E2b-deleted and gutted vectors
FT      CC provide improved adenoviral vectors useful for a wide variety of
FT      CC gene therapy applications.
FT      SQ Sequence 34303 BP; 7948 A; 9625 C; 9517 G; 7213 T;
Query Match 100.0%; Score 345; DB 1; Length 34303;
Best Local Similarity 100.0%; Pred. No. 1.4e-104;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgtttcttcagctcttcctccgctctccgctgtgtgtgactgcagaaagatgtgtagt 60
DB 32711 ATGTTCTTCAGAGCTCTCCGCTCTCCGCTGTGACTGCAGAAAGATGTAGGT 32652
QY 61 tggctgggtgtggtctatcttcggtgtgtgtgtatcatcagggcggcgcatgaaga 120
DB 32651 TGGCTGGGTGTGCTTATCTTGGGTGTGATCATCAGGGCGCGCGCATGAAGA 32592
QY 121 gttacatagaaaccgaagccagggcgccctggtgtgtttacagagtgatatactac 180
DB 32591 GTTACATAGAAACCGAAGCCAGGGCGCGCTGTGATGCTTTCAGAGATGATATAC 32532
QY 181 aactactacagagcgatctaaagcgagagccgagagcgagatctgttttcacg 240
DB 32531 AACTACTACAGAGCGATCTTAAGCGCGAGACGCGAGACGAGATCTGTTCACGCC 32472
QY 241 cgacactgtgtttgtttccttaagaaatatactacgtccggttcattgtgatac 300
DB 32471 CGCACCTGTGTTTGTTCAGAAATATATACATGATCGCGCTTCATTGGGATGAC 32412
QY 301 acgacacacacgatctcgtgtctcggcgacactcgtacagtag 345
DB 32411 ACGACACACGATCTCGGTGTCTCGCGGACACTCCGATACAGTAG 32367

```

```

RESULT 6
ID X15627/C
AC X15627;
DC 07-MAY-1999 (first entry)
DE Recombinant adenovirus Ad:Pac-beta-Gal; replication deficient adenovirus;
KW E1a region: E3 region; gene therapy; nitric oxide synthetase; NOS;
KW cystic fibrosis chloride channel; LDL receptor; erythropoietin;
KW atherosclerotic artery; ss.
OS Synthetic.
OS Mastadenovirus.
FH Key location/Qualifiers
FT CDS complement (3372..333)
FT      /*tag= a
FT      US5880102-A.
FT      09-MAR-1999.
FT      17-JAN-1995; 374483.
FT      17-JAN-1995; US-374483.
FT      (UYDU-) UNIV DUKE.
FT      Blazing NA, George SE;
FT      WPI: 99-204005/17.
FT      P-PSDB: W97243.
FT      New replication deficient adenovirus bearing deletions of the E1a
FT      and E3 regions - containing a single packaging signal sequence and
FT      E1a enhancer sequence, the E1a deletion has unique cleavage sites
FT      and is useful as a gene therapy vector.
FT      Disclosure: Fig 49-100, 148pp; English.
FT      CC The present sequence represents recombinant adenovirus
FT      CC Ad:Pac-beta-Gal, which exemplifies the invention. The specification
FT      CC describes replication deficient adenoviruses having deletions of
FT      CC the E1a and E3 regions and comprise a single packaging signal sequence
FT      CC and E1a enhancer sequence, where the sequences are at the 3' end of the
FT      CC adenovirus and the E1a deletion contains at least one PacI, ClaI, XbaI
FT      CC or BstBI cleavage site. The replication deficient viral vectors can be
FT      CC used in gene therapy regimens to effect the transfer of genes encoding
FT      CC molecules of therapeutic importance, including isoforms of the nitric
FT      CC oxide synthetase (NOS) gene (brain, endothelial and microphage NOS), the
FT      CC cystic fibrosis chloride channel (CFTR) gene, the dystrophin gene, the
FT      CC LDL receptor gene and the erythropoietin gene. The NOS isoforms can be
FT      CC used in vascular applications or in cancer therapy (microphage NOS). The
FT      CC NOS gene can be introduced into vein grafts prior to their use as
FT      CC coronary artery bypass grafts. A NOS containing adenovirus can also be
FT      CC used following coronary angioplasty to prevent retinosis and to treat
FT      CC atherosclerotic arteries.
FT      SQ Sequence 34382 BP; 7923 A; 9880 C; 9421 G; 7158 T;
Query Match 100.0%; Score 345; DB 1; Length 34382;
Best Local Similarity 100.0%; Pred. No. 1.4e-104;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgtttcttcagctcttcctccgctctccgctgtgtgtgactgcagaaagatgtgtagt 60
DB 32543 ATGTTCTTCAGAGCTCTCCGCTCTCCGCTGTGACTGCAGAAAGATGTAGGT 32484
QY 61 tggctgggtgtggtctatcttcggtgtgtgtgtatcatcagggcggcgcatgaaga 120
DB 32483 TGGCTGGGTGTGCTTATCTTGGGTGTGATCATCAGGGCGCGCGCATGAAGA 32424
QY 121 gttacatagaaaccgaagccagggcgccctggtgtgtttacagagtgatatactac 180
DB 32423 GTTACATAGAAACCGAAGCCAGGGCGCGCTGTGATGCTTTCAGAGATGATATAC 32364
QY 181 aactactacagagcgatctaaagcgagagccgagagcgagatctgttttcacg 240
DB 32363 AACTACTACAGAGCGATCTTAAGCGCGAGACGCGAGACGAGATCTGTTCACGCC 32304
QY 241 cgacactgtgtttgtttccttaagaaatatactacgtccggttcattgtgatac 300
DB 32303 CGCACCTGTGTTTGTTCAGAAATATATACATGATCGCGCTTCATTGGGATGAC 32244

```



```

DB 34162 AACCTACTACACAGACGAGATCTAGACGGGACAGACCGGAGACGACGAGATCTGTTGTACAGCC 34103
OY 241 GCGACCTGTTTGGTCTGGGAAATATGCTACCTCTCCGCGCTTCCATTTGGATGATGACT 300
DB 34102 GCGACCTGTTTGGTCTCGGAAATATGCTACCTACGTCGCCGCTTCCATTTGGCATGACACT 34043
OY 301 acgacacacacgacatctcgattgtctcggcgacactccgtacagtag 345
DB 34042 ACGACCAACACAGATCTCGGTTGTCGCGGACACTCGGTACAGTAG 33998
RESULT
ID T60558/c 9
ID T60558: standard; DNA; 36538 BP.
AC T60558:
AT 05-MAR-1998 (first entry)
DE Recombinant trans-acting adenovirus H5 020CMVag-pol.
KM Adenovirus H5 020CMVag-pol; plasmid pAdMLVneo-int; Ad5; Mo-MLV;
NS retrovirus; retrotransposition; transposition; transgene;
OS gene therapy; vector; ss.
OS Chimeric - Mastadenovirus serotype 5.
OS Chimeric - Moloney murine leukaemia virus.
OS Chimeric - Cytomegalovirus.
PM W09715679-AL.
PN 01-MAR-1997.
PD 24-OCT-1996; 017176.
PE 27-OCT-1995; US-005942.
PR (TYPE ) UNIV PENNSYLVANIA.
PI Kelley WM, Wilson JM;
PI WPI: 97-259031/23.
DR Recombinant replication defective virus - inserts transgene into
PT host cell chromatin in the presence of transposase, providing stable
PT and durable transgene expression
PS Example 2; Fig 9A-1; 74pp; English.
SC This is the DNA sequence of the trans-acting recombinant adenovirus
CC H5 020CMVag-pol. It was obtained by co-transfecting HEK293
CC (ANC Cris1573) cells with linearised plasmid pAdCMVag-pol (see
CC T60555) and Ad5d17001. A cis-acting adenovirus (see T60557) was
CC also produced, and retrotransposition and integration into host
CC cells was demonstrated in HeLa cells infected with both viruses.
CC Claimed novel recombinant replication defective viruses (A)
CC comprise: (a) DNA of, or corresponding to, at least part of the
CC viral genome, able to infect a mammalian cell; and (b) a first
CC expression sequence (ES1) comprising a human gene (I) linked to
CC regulatory sequences for expression, with the cassette flanked by
CC the cis-acting terminal repeat sequences of a transposon, and ES1
CC flanked by DNA of (a); (A) can infect a mammalian cell and can
CC express (I) and transfer it to the cellular chromatin in vivo or in
CC vitro in the presence of a transposase. Also new are: (1) a
CC recombinant replication defective virus (B) containing (a) as above
CC and a second expression sequence (ES2) containing a trans-acting
CC transposase gene (II) plus regulatory sequences, flanked by DNA of
CC (a) and able to express transposase in mammalian cells; and (2)
CC mammalian cells which stably express (I) integrated into its
CC chromatin, produced by infection with (A). (A) and (B) are used
CC for delivery and stable integration of (I) into a host cell
CC (claimed), e.g. in somatic gene therapy of genetic defects or
CC deficiencies such as cystic fibrosis. (A) are also used to prepare
CC recombinant retroviruses (RV) (claimed). Because (I) becomes
CC stably integrated, it provides longer lasting expression than genes
CC introduced with conventional adenoviral vectors and the need for
CC repeated administration is avoided. The transgene is inherited by
CC progeny cells and the viral component is gradually degraded by the
CC cell. When used for RV production, (A) increase the ratio of RV
CC to empty retrovirus.
SQ Sequence 36538 BP; 8621 A; 10490 C; 9909 G; 7518 T;
Query Match 100.0%; Score 345; DB 1; Length 36538;
Best Local Similarity 100.0%; Pred. No. 1,5e-104;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0
1 atggtcttcacgctcttcgcgcctcccgctggtgactcgcgaagaacgaatgtagt 60

```

Db	34945	ATGGTCTCTCCAGCCTCTTCCCGCTCCTCCCGTGTGACTCCGCAACGAAATGTGNAGT	34886
Qy	61	tgccgtgggtgtgtgcttattcttcgcgtgtgtgatagtttatcaaggcagcgcgcatgaagta	120
Db	34885	TGGCGGGGtGTGGCTATTCTTCCTGCGGTGTGGATATGATATCAGGCGAGCGCGCATGAAGTA	34826
Qy	121	gtttacatgaagaccgaagccagcgaggcgccctgcgaatgcctttagagagtgatatactac	180
Db	34825	GTTTTCATATGAAGACCCGGAAGCCAGGGGGCGCCCTTGATGCTTTAGAGAGTGTATATCTAC	34766
Qy	181	aactactacacagagcgatctaaagcgcgagacacggagagacagcatctgttgcaagcc	240
Db	34765	AACTCTCTACACAGACGATCTAAGCGGGGAGACCCGAGACCGAGATCTGTTGTACCGCC	34706
Qy	241	ggcaccctgatttggttcaaggaatactactacgtcccgagcttccatttggatgacct	300
Db	34705	CGCAGCTGGTTTGTGCTTCAAGAAATATGACTACGTCCGGCGCTTCCATTGGCATGACACT	34646
Qy	301	acgaccacacacagatctcggtgtgtctcggcgacatccgtacagtag	345
Db	34645	ACGACCAACACGATCTCGGTTGTCTCGCGGCACTCCGTACAGTAG	34601
RESULT	10		
ID	X07371/c	X07371 standard; DNA: 34427 BP.	
AC	X07371;		
DT	07-JUN-1999	(first entry)	
DE	Adenovirus vector 231-10 genome.		
KM	Adenovirus; vector 231-10; apoptosis; inhibitor; R1D-alpha;		
KM	R1D-beta; receptor internalisation and degradation; cancer;		
KM	degenerative disease; immune disorder; gene therapy;		
KM	tissue transplant; death receptor; tumour necrosis factor receptor;		
KM	human; ds.		
OS	Chimeric - Mastadenovirus 5.		
OS	Chimeric - Mastadenovirus 2.		
OS	Chimeric - Human cytomegalovirus.		
PN	WO9002658-A1.		
PD	21-JAN-1999.		
PF	08-JUL-1998; U14239.		
PR	09-JUL-1997; US-088939.		
PA	(UWSL-) UNIV SAINT LOUIS.		
PI	W01d WSM:		
DR	WPI: 99-120862/10.		
PT	Method for inhibiting apoptosis - using receptor internalisation and		
FT	degradation (R1D) complex		
PS	Example 10; Fig 28A-XX: 126pp; English.		
CC	This is the nucleotide sequence of vector 231-10, a human		
CC	adenovirus serotype 5 (Ad5) vector. 231-10 lacks the E1A, E1B		
CC	and E3 transcription units. The deleted E1A and E1B regions are		
CC	replaced with an expression cassette wherein all the E3 proteins		
CC	are expressed from the human cytomegalovirus (CMV) promoter. The		
CC	E3 expression cassette contains the E3 genes from virus pm734.1,		
CC	a derivative of virus rec700, which is an Ad5-Ad2-Ad recombinant		
CC	that has the Ad2 version of the E3 genes for the 12.5k, 6.7k,		
CC	gpl9k and R1D-alpha proteins, and the Ad5 version of the E3 genes		
CC	for the R1D-beta and 14.7k proteins. In addition, the vector has		
CC	2 missense mutations in the adp (adenovirus death protein) gene		
CC	that eliminate the first 2 met codons, thereby precluding synthesis		
CC	of functional ADP. Because 231-10 lacks E1A, viral genes in the		
CC	vector backbone are not expressed; only the E3 proteins are		
CC	expressed from the CMV promoter. Thus, the vector serves as an		
CC	essentially inert vehicle by which E3 genes can be delivered into		
CC	cells. The invention provides a method for inhibiting apoptosis of		
CC	a cell expressing a death receptor of the tumour necrosis factor		
CC	receptor family. This involves treating the cells with a receptor		
CC	internalisation and degradation (R1D) protein complex containing		
CC	R1D-alpha (10.4 k, see W97877 and W97878) and R1D-beta (14.5k, see		
CC	W97879) proteins encoded by the E3 region of adenovirus. The cell		
CC	can be treated by administering to the cell a polynucleotide		
CC	expressing the R1D complex or a composition containing the R1D		
CC	complex. The compositions and methods are used to treat		

[illegible]

CC degenerative immunodeficiency diseases, particularly to decrease
 CC leukocyte apoptosis (claimed). Immune disorders that can be
 CC treated include autoimmune disorders, ischemic injury such as
 CC caused by myocardial infarction, stroke induced neuron death and
 CC reperfusion injury, alcohol-induced hepatitis, diseases caused by
 CC viral infection such as AIDS and fulminant hepatitis, and cancer.
 CC The methods can also be used to promote tissue transplant
 CC survival. Vector 231-10 can be used to deliver the RID complex to
 CC the cells.
 SQ Sequence 34427 BP; 8098 A; 9606 C; 9451 G; 7272 T;

Query Match 99.5%; Score 343.4; DB 1; Length 34427;
 Best Local Similarity 99.7%; Pred. No. 4.9e-104;
 Matches 344; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 atggtcttcacagctctccgcgtccgcgtgtgtgactgcagacgaatgtgtagt 60
 DB 32834 ATGGTCTTCACAGCTCTCCGCTCCCTCCGCTGTGTGACTGCAGAACGATGTGTAGT 32775
 QY 61 tggctgggtgtggtctatctcgtggtgtgtgtatcatcaggcgagcgagcagaaagga 120
 DB 32774 TGGCTGGGTGTGGCTTATCTGCGGTGTGTATCATAGGCGAGCGCCGCTATGAGGA 32715
 QY 121 gttacatagaaacccagagcagagcgagcgctgtgactgtcttgaagagtgatatactac 180
 DB 32714 GTTATACATAGAACCCAGAGCCAGAGCGGCGCTGTGTGAGAGATGATATATAC 32655
 QY 181 aactactacacagagcagatctaaagcgagagacgagacgagatctgttgcagcgc 240
 DB 32654 AACTACTACACAGAGCAGATCTAAAGCGAGAGCCGAGACGAGATCTGTGTACGCGC 32595
 QY 241 cgcacctgtgtgttgaagaataatgactacgtcgcgcgttcattgtgcatgacct 300
 DB 32554 CGCACCCTGTGTGCTTATGAGAAATATGACTACGTCCGCGCTTCATTGGCATACACT 32555
 QY 301 acgaccacacagatctcgtgtgtcgcgcgcacatcgcgtacagtag 345
 DB 32534 ACGACCACACAGATCTCGGTGTCTCGGCGCAGCTCCGTACAGTAG 32490

RESULT 11
 V22141/c
 ID V22141 standard; cDNA; 36519 BP.
 AC V22141
 DT 17-AUG-1998 (first entry)
 DE Chimpanzee adenovirus C68 genomic sequence.
 KW genomic sequence; viral genome; virus; gene therapy; treatment;
 KW acquired defects; inherited defects; genetic engineering; vector;
 KW in vitro production; recombinant protein; ds.
 OS Chimpanzee adenovirus C68.
 FT Key Location/Qualifiers
 FT CDS 480..1521
 FT /tag- a
 FT /note- "E1A gene"
 FT 1560..3956
 FT /tag- b
 FT CDS /note- "E1B gene"
 FT complement(23370..21787)
 FT /tag- c
 FT /note- "E2A gene"
 FT complement(10346..3957)
 FT /tag- d
 FT CDS /note- "E2B gene"
 FT 26806..31877
 FT /tag- e
 FT /note- "E3 gene"
 FT complement(36193..33486)
 FT /tag- f
 FT CDS /note- "E4 gene"
 FT 10823..13817
 FT /tag- g
 FT CDS /note- "L1 gene"

FT CDS 13884..17431
 FT /tag- h
 FT /note- "L2 gene"
 FT 17480..21804
 FT /tag- i
 FT CDS /note- "L3 gene"
 FT 23399..27439
 FT /tag- j
 FT /note- "L4 gene"
 FT 32134..33502
 FT /tag- k
 FT CDS /note- "L5 gene"
 FT /note- "L5 gene"
 FT PN WO9810087-A1.
 FT PD 12-MAR-1998.
 FT PF 04-SEP-1997; U15694.
 FT PR 06-SEP-1996; US-024700.
 FT PA (UYPE-) UNIV PENNSYLVANIA.
 FT PI Farina SF, Fisher KJ, Wilson JM;
 FT WP: 98-193635/17.
 PT Chimpanzee adenovirus vector - useful in gene therapy and genetic
 PT engineering in general
 PS Disclosure: Pages 70-89; 116pp; English.
 CC The sequence is that of the Chimpanzee adenovirus C68. It may be
 CC used in the production of a vector comprising it and a selected
 CC heterologous gene operatively linked to regulatory sequences
 CC directing its expression in a heterologous cell.
 CC Such recombinant adenoviruses are useful as vectors in gene
 CC therapy and genetic engineering in general, especially for
 CC the treatment of acquired or genetically inherited defects.
 CC The viruses are also useful for the in vitro production of
 CC recombinant proteins of interest.
 SQ Sequence 36519 BP; 7809 A; 10790 C; 10623 G; 7297 T;

Query Match 41.7%; Score 143.8; DB 1; Length 36519;
 Best Local Similarity 64.0%; Pred. No. 1e-37;
 Matches 217; Conservative 0; Mismatches 122; Indels 0; Gaps 0;
 QY 1 atggtcttcacagctctccgcgtccgcgtgtgtgactgcagacgaatgtgtagt 60
 DB 34942 ATGGTCTTCCTGTTCTTCCCTCCCGCCGCTCAACGAAACCAAAATGATCATC 34883
 QY 61 tggctgggtgtggtctatctcgtggtgtgtgtatcatcaggcgagcgagcagaaagga 120
 DB 34882 TGGCTGGGTGTGGCTTATCTGCGGTGTGTATCATAGGCGAGCGCCGCTATGAGGA 34823
 QY 121 gttacatagaaacccagagcagagcgagcgctgtgactgttgaagagtgatatactac 180
 DB 34822 ATCTTATACACTCCGAGAGCCCTAGATGTGTGACAGGCTGAGAGAGTGTGTTAC 34763
 QY 181 aactactacacagagcagatctaaagcgagagacgagacgagatctgttgcagcgc 240
 DB 34762 AATTTCACACAGCAGCTTCCCAAGCGCAGACCGCCGCGATCGGTGTGACGCC 34703
 QY 241 cgcacctgtgtgttgaagaataatgactacgtcgcgcgttcattgtgcatgacct 300
 DB 34702 CGGACCAAGTCTGCTAAGCAAGATGCAAAATGTCCAAACCAATCATCATGACAG 34643
 QY 301 acgaccacacagatctcgtgtgtcgcgcgcacatcgcgtacagtag 339
 DB 34642 GTCCGGAACACAGATACGCCGCTCCCGCATCACCCGTA 34604

RESULT 12
 V22140/c
 ID V22140 standard; cDNA; 35524 BP.
 AC V22140
 DT 17-AUG-1998 (first entry)
 DE Chimpanzee adenovirus C1 genomic sequence.
 KW genomic sequence; viral genome; virus; gene therapy; treatment;
 KW acquired defects; inherited defects; genetic engineering; vector;
 KW in vitro production; recombinant protein; ds.
 OS Chimpanzee adenovirus C1.

	Key	Location/Qualifiers
FH	CDS	480..1540
FT		/tag- a
FT		/note- "E1A gene"
FT		1566..3958
FT	CDS	/tag- b
FT		/note- "E1B gene"
FT		complement(23665..22065)
FT	CDS	/tag- c
FT		/note- "E2A gene"
FT		complement(10379..3959)
FT		/tag- d
FT		/note- "E2B gene"
FT		27181..31375
FT	CDS	/tag- e
FT		/note- "E3 gene"
FT		complement(35228..32535)
FT		/tag- f
FT		/note- "E4 gene"
FT		10893..13864
FT	CDS	/tag- g
FT		/note- "L1 gene"
FT		13925..17591
FT		/tag- h
FT	CDS	/note- "L2 gene"
FT		17641..22083
FT		/tag- i
FT		/note- "L3 gene"
FT		23697..27813
FT		/tag- j
FT		/note- "L4 gene"
FT		31556..32551
FT		/tag- k
FT		/note- "L5 gene"
PN	WO9810087-A1.	
PD	12-MAR-1998.	
PF	04-SEP-1997:	U15694.
PR	06-SEP-1996:	US-024700.
PA	(TYPE-) UNIV PENNSYLVANIA.	
PI	Farina SF, Fisher KJ, Wilson JM;	
PT	WPI; 96-193635/17.	
PS	Chimpanzee adenovirus vector - useful in gene therapy and genetic engineering in general	
PP	Disclosure: Pages 51-70; 116pp; English.	
CC	The sequence is that of the chimpanzee adenovirus C1. It may be used in the production of a vector comprising it and a selected heterologous gene operatively linked to regulatory sequences	
CC	directing its expression in a heterologous cell.	
CC	such recombinant adenoviruses are useful as vectors in gene therapy and genetic engineering in general, especially for the treatment of acquired or genetically inherited defects.	
CC	The viruses are also useful for the in vitro production of recombinant proteins of interest.	
SQ	Sequence 35524 BP; 8873 A; 9142 C; 9033 G; 8476 T;	
Query Match	29.9%; Score 103; DB 1; Length 35524;	
Best Local Similarity	60.6%; Pred. No. 3.6e-24;	
Matches 169; Conservative 0; Mismatches 110; Indels 0; Gaps 0;		
Oy	1 atggtcttcacagcgccttccgcgtctccgcgtgctgactcgacagaacatgctagtgt 60 	
Dd	33994 ATGGTCTTCACAGTCCTTCCTCCCTCCTTTGATTAAACTCAAGCTAATGTATGCCA 33935 	
Oy	61 tggcttggtgtggtattctcggtgtgtgatagttatcaaggcagcgcgcatgaagaa 120 	
Dd	33934 TGCGTGCGGATGGCCCCATTCCTACAAGTAATGATTTATACAGCCGATCAGACATGATGGA 33875 	
Oy	121 gtccatcagaaccgccagggcgcgctcgatgctctttagagagtgatatactac 180 	
Dd	33874 GTTTCATTAACAACTGAAGCCGAGAAGCTTTTGACCACCAACTCAGGGGAATGCCTGTACTTC 33815 	
Oy	181 nactactcacagagcgatctcaagcgcgcgagacccggagacgcagatctgtttgttaagcc 240 	

Db	33814	AATTACCACTGAATGCGTTAGGGCAGCGGATPAGAAGCCGAAGAAGTGTGTGTGCT	33755
Oy	241	cgcactggttttcctcaggaatatgctcgccgg	279
Db	33754	AMGCTAGGTTTGTCATGCAAAATAACGAAATGTCAGG	33716
 RESULT 13 068003/c			
ID	068003	standard; DNA: 36335 BP.	
AC	068003;		
DT	26-MAR-1996	(first entry)	
DE	Ad2/-ORF6/PgK-CFR nucleotide sequence.		
KW	Recombinant adenovirus; Ad2/CFR-1; adenovirus 2 serotype; Ela; E1b;		
RX	Viral replication; gene expression; gene therapy; cystic fibrosis;		
KW	Cystic fibrosis transmembrane conductance regulator; CFR;		
KM	Promoter; E3; p13; MHC; class I; viral latency; pulmonary airway; ds.		
OS	Synthetic.		
FT	Key	Location/Qualifiers	
FT	misc_feature	12915..36335	
FT	/tag= a	/note= "Represents residues 10676-34096 of Ad2-E4/ORF6"	
FT	35069..35973		
FT	/tag= b	/note= "Represents residues 33178-34082 of Ad2"	
FT	12915..35054		
FT	/tag= c	/note= "Represents residues 1-32815of Ad2"	
FT	28478..28790		
FT	/tag= d	/number= 1	
FT	/note= "33k protein"	28791..28992	
FT	/tag= e	/number= 1	
FT	28993..29366		
FT	/tag= f	/number= 2	
FT	/note= "33k protein"	13279..14526	
FT	/tag= g	/product= 52,55k protein	
FT	14547..16304		
FT	/tag= h	/product= IIIta protein	
FT	/note= "Peripentonal hexon-associated protein"	16331..16336	
FT	/tag= i	/number= 1	
FT	/note= "Major late mRNA L1 poly A signal (putative)"	16390..18105	
FT	/tag= j	/product= Penton protein	
FT	/note= "Viron component III"	18112..18708	
FT	/tag= k	/product= Pro-VII protein	
FT	/note= "Precursor to major core protein"	18778..19887	
FT	/tag= l	/product= PV protein	
FT	/note= "minor core protein"	20188..20193	
FT	/tag= m	/note= "major late mRNA L2 poly-A signal (putative)"	
FT	20240..20992		
FT	/tag= n	/product= pVI protein	
FT	/note= "Hexon associated precursor"	21077..23983	
FT	/tag= o	/product= Hexon protein	
FT	/note= "Viron component II"		

FT	polya_signal	24657..24662	/*tag- p
FT		/note= "Major late mRNA l3 poly-A signal (putative)"	
FT	cds	complement (244729..26318)	
FT		/*tag- q	
FT		/product= DBP protein	
FT		/note= "DNA binding or 72k protein"	
FT	cds	26347..28764	
FT		/*tag- i	
FT		/product= 100K protein	
FT		/note= "Hexon assembled"	
FT		29454..30137	
FT		/*tag- s	
FT		/product= pVII protein	
FT	polya_signal	/note= "Hexon-associated precursor"	
FT		30444..30449	
FT		/*tag- t	
FT	cds	/note= "major late mRNA l4 poly-A signal (putative)"	
FT		31051..31350	
FT		/*tag- u	
FT		/product= E3 19K protein	
FT	cds	/note= "Glycosylated membrane protein"	
FT		31707..32012	
FT		/*tag- v	
FT		/product= E3 11.6K protein	
FT	polya_signal	32008..32013	
FT		/*tag- w	
FT		/note= "E3-1 mRNA poly-A signal (putative)"	
FT	polya_signal	33081..33086	
FT		/*tag- x	
FT		/note= "E3-2 mRNA poly-A signal (putative)"	
FT	polya_signal	35015..35018	
FT		/*tag- y	
FT	misc_feature	/note= "major late mRNA l5 poly-A signal (putative)"	
FT		1..12914	
FT		/*tag- z	
FT		/note= "Represents residues 1-12914 of pAd2/Pgk-CpTR"	
FT	promoter	380..914	
FT		/*tag- aa	
FT		/note= "pgk promoter"	
FT	cds	1011..5453	
FT		/*tag- ab	
FT		/product= CpTR	
PD	W09125649-A.		
PD	09-JUN-1994.		
PF	02-DEC-1993.	U11667.	
PR	03-DEC-1993.	US-985478.	
PR	01-OCT-1993.	US-130682.	
PR	13-OCT-1993.	US-136742.	
PA	(GENZ) GENZYME CORP.		
PI	Armentano D, Couture LA, Gregory RJ, Smith AE;		
PI	WPI: 94-200277/24.		
PT	Adeno:virus-based gene therapy vectors - esp. useful for gene		
PT	therapy of cystic fibrosis		
PS	Example 15; Page 84-95; 167pp; English.		
CC	This sequence represents the nucleotide sequence of the second generation		
CC	adenovirus vector named Ad2-ORF6/Pgk-CpTR. This virus lacks E1 and in its		
CC	place contains a modified transcription unit with the phosphoglycerate		
CC	kinase (Pgk) promoter and a poly A addition site flanking the cystic		
CC	fibrosis transmembrane conductance regulator (CFTR) cDNA. The Pgk		
CC	promoter is only if moderate strength but it is long lasting and is not		
CC	subject to shut off. The E4 region of the vector has also been modified		
CC	in that the whole sequence has been removed and replace by ORF6, the		
CC	only E4 gene essential for growth of adenovirus (Ad) in tissue culture.		
CC	The DNA construct comprises a full length copy of the Ad2 genome from		
CC	which the early region 1 genes (E1 genes) have been deleted and replaced		
CC	by an expression cassette encoding CFTR. The expression cassette		
CC	includes the promoter for Pgk and a poly-A addition signal from the		
CC	bovine growth hormone gene. The Ad2-ORF6/Pgk-CpTR construct differs		
CC	from that given in G68002 (Ad2/CFTR-1). In that the latter utilizes the		
CC	endogenous E1a promoter, has no poly-A addition signal directly		
CC	downstream of CFTR and retains an intact E4 region. This adenovirus may		
CC	be administered to the pulmonary airways in the gene therapy of cystic		

```

CC      fibrosis
SQ      Sequence      36335 BP;      8597 A;      10000 C;      9786 G;      7952 T;

Query Match      23.2%;      Score 80;      DB 1;      Length 36335;
Best Local Similarity      98.9%;      Pred. No. 2.5e-18;
Matches      86;      Conservative      0;      Mismatches      1;      Indels      0;      Gaps      0;

OY      259      aggaatatgactagctcgcgcgttcacattcggcactgacacacacacagatctcg      318
Db      35975      ATGAAATATGACTACTAGCTCGCGCGTTCCTTTGGCATGACTACACACCAACAGATCTCG      35916
OY      319      gtgtctcgcgcgcactccgctacagtag      345
Db      35915      GTTGCTCGCGCACCTCCGTACAGTAG      35889

RESULT      14
T59270
ID      T59270 standard;      CDNA;      3653 BP.
AC      T59270;
DT      27-AUG-1997 (first entry)
DE      Plasmid pMW7E4ORF6 minigene sequence.
KW      Minigene; human adenovirus type 5; open reading frame; promoter; in vivo;
KW      mouse mammary tumour virus; growth hormone; terminator; recombinant;
KW      packaging cell line; adeno-associated virus; ex vivo; gene therapy;
KW      inherited disease; cancer; genetic dysfunction; cystic fibrosis; ds.
OS      Synthetic.
FH      Key
FT      Location/Qualifiers
FT      1..1506
FT      /tag= a
FT      /note= "mouse mammary tumour virus promoter"
FT      cds
FT      1521..2405
FT      /tag= b
FT      /product= human Ad5 E4 region ORF6 protein
FT      /note= "specification gives this CDS at nucleotides
FT      1523-2408"
FT      /transl_except= (pos: 1863..1865, aa: Tyr)
FT      /transl_except= (pos: 2091..2093, aa: Tyr)
FT      2409..3654
FT      /tag= c
FT      /note= "growth hormone gene terminator sequence"

terminator
M09639530-A2.
PD      12-DEC-1986.
PF      04-JUN-1986;      U10245.
PR      05-JUN-1995;      US-462014.
PR      27-OCT-1995;      US-549489.
PA      (UYPE) UNIV PENNSYLVANIA.
PI      Fisher KJ, Gao G, Wilson JM;
DR      WPI: 97-0043152/04.
DR      P-PSDB: W01729.
PT      Recombinant adeno-associated virus contg. second gene which
PT      facilitates its conversion from single stranded to double stranded
PS      virus - enhances efficiency of ex vivo transduction into target cell
PS      Example 1: Page 69-72; 131pp; English.
CC      This is the nucleotide sequence of a minigene containing the human
CC      adenovirus type 5 (Ad5) E4 region open reading frame (ORF) 6 under
CC      control of the mouse mammary tumour virus promoter and a growth hormone
CC      terminator sequence. The minigene is used to generate a packaging cell
CC      line expressing the E4 ORF6 product which is used to rescue and package
CC      an E4 deficient recombinant adeno-associated virus (rAAV). The rAAV
CC      comprises at least 2 selected genes; the first gene able to express the
CC      gene of interest in the target cell; and the second gene able to convert
CC      the single stranded virus to its double stranded form upon expression.
CC      The rAAV is useful in pharmaceutical compositions for use in ex vivo and
CC      in vivo gene therapy for the treatment of inherited diseases, cancer and
CC      other genetic dysfunctions, e.g. cystic fibrosis.
SQ      Sequence      3653 BP;      860 A;      851 C;      933 G;      963 T;

```

QY 266 atgactacgtccggcgttcattggcagatgacacacacacagatcgtgtct 325
DB 1521 ATGACTACGTCGGCGGCTTCATTGGCATGACACTACGACCAACAGATCTGGTGTCT 1580
QY 326 cggcgacactccgtacagtag 345
DB 1581 CGCGCACTCCGCTACAGTAG 1600

RESULT 15

T59273

ID T59273 standard; DNA; 8299 BP.

AC T59273;

DE 27-AUG-1997 (first entry)

DB Second generation recombinant adenoviral vector pAV.CMVLP.GRE-ORF6.

TX Mouse mammary tumour virus; growth hormone; terminator; recombinant;

TX Packaging cell line; adeno-associated virus; ex vivo; gene therapy;

TX Inherited disease; cancer; genetic dysfunction; cystic fibrosis; ds.

OS Synthetic

FH Key

FH repeat_unit

FH

FH

FH promoter

FH

FH cds

FH

FH poly_a_signal

FH

FH promoter

FH

FH cds

FH

FH poly_a_signal

FH

FH repeat_unit

FH

FH misc_feature

FH

FH

FH

FH

FH

FH

FH

FH

FH

FH

FH

FH

FH

FH

FH

FH

FH

FH

FH

FH

FH

FH

FH

CC virus.
SQ Sequence 8299 BP; 1816 A; 2130 C; 2108 G; 1963 T;
Query Match 23.2%; Score 80; DB 1; Length 8299;
Best Local Similarity 100.0%; Pred. No. 8.7e-17;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 266 atgactacgtccggcgttcattggcagatgacacacacacagatcgtgtct 325
DB 3402 ATGACTACGTCGGCGGCTTCATTGGCATGACACTACGACCAACAGATCTGGTGTCT 3461
QY 326 cggcgacactccgtacagtag 345
DB 3462 CGCGCACTCCGCTACAGTAG 3481

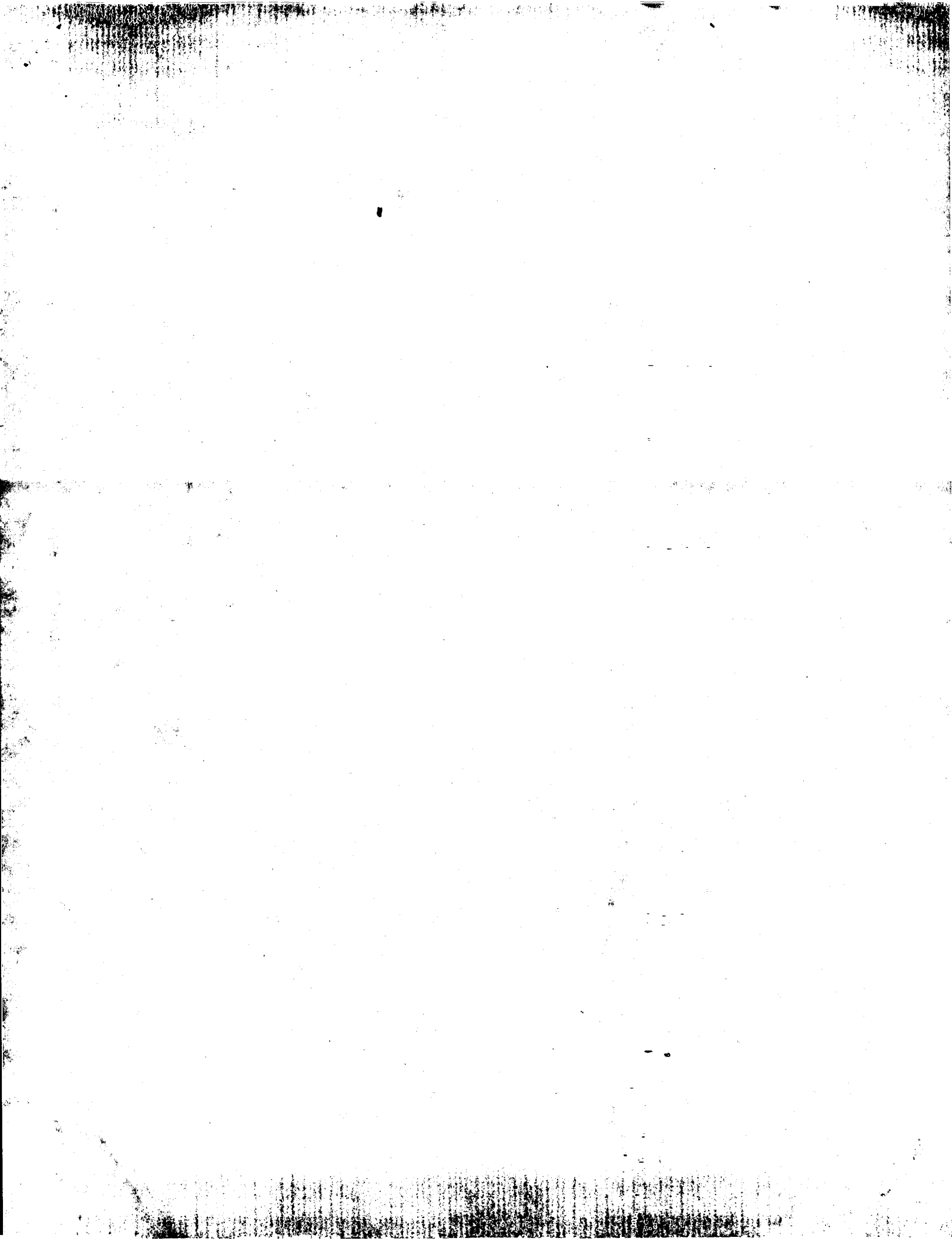
Search completed: June 23, 2000, 10:06:58
Job time: 3686 sec

Location/Qualifiers
53.219
/tag- a
/rpt_type- inverted
/note- "5' AAV ITR"
255.848
/tag- b
/function- enhancer promoter
/note- "cytomegalovirus enhancer/promoter"
914.2892
/tag- c
/product- human placental alkaline phosphatase
2893.3090
/tag- d
/note- "SV40 poly-A sequence"
314.3393
/tag- e
/note- "glucocorticoid dependent promoter GRE"
3402.4286
/tag- f
/product- Ad5 E4 ORF6 product.
4315.4713
/tag- g
/note- "SV40 poly-A sequence"
4547.4713
/tag- h
/rpt_type- inverted
/note- "3' AAV ITR"
4714.8299
/tag- i
/note- "plasmid derived sequences"
W09639330-A2.
12-DEC-1996.
04-JUN-1996; U10245.
05-JUN-1995; US-462014.
27-OCT-1995; US-549489.
PA (TYPE-) UNIV PENNSYLVANIA.
PI Fisher KJ, Gao G, Wilson JM;
WPI: 97-043152/04.
DR Recombinant adeno-associated virus contg. second gene which
FT facilitates its conversion from single stranded to double stranded
PT virus - enhances efficiency of ex vivo transduction into target cell
PS Example 19; Page 108-114; 131pp; English.
CC This is the nucleotide sequence of a second generation of recombinant
CC adeno-associated virus (rAAV) which contains 2 genes, the first gene
CC able to express the gene of interest (beta-galactosidase) in the target
CC cell, and the second gene(Ad5 E4 ORF6) able to convert the single
CC stranded virus to its double stranded form upon expression. The rAAV is
CC useful in pharmaceutical compositions for use in ex vivo and in vivo gene
CC therapy for the treatment of inherited diseases, cancer and other genetic
CC dysfunctions, e.g. cystic fibrosis.
CC See also T59271 for another example of a recombinant adeno-associated

Mon Jun 26 09:10:45 2000

us-09-214-478-3.rng

Page 11



```

RESULT      1
LOCUS       A52460/c
DEFINITION  Sequence 4 from Patent WO9622378.
ACCESSION   A52460
VERSION     A52460.1 GI:2851930
KEYWORDS    .
SOURCE      unidentified.
ORGANISM    unidentified
            unclassified.
REFERENCE   1 (bases 1 to 3189)
AUTHORS     Dedieu,J., Latte,M., Orsini,C., Perricaudet,M., Vigne,E. and Yeh,P.
TITLE       CELLS FOR THE PRODUCTION OF RECOMBINANT ADENOVIRUSES
JOURNAL     PATENT: NO 9522378-A 4 25-JUL-1996;
            RHONE POULENC ROGER SA (FR)
OTHER PUBLICATION AV 4544396 960807
            Other publication FR 2729674 960726.
FEATURES             Location/Qualifiers
     source          1..3189
                     /organism="unidentified"
                     /db_xref="taxon:32644"
BASE-COUNT        975 a           912 c           631 g           671 t
ORIGIN

```

Query Match	100.0%	Score 345	DB 5	length 3189
Best Local Similarity	100.0%	Pred. No. 5	5e-97	
Matches 345	Conservative	0	Mismatches 0	Indels 0
				Gaps 0

Qy 1 atggttcttcacagctcttcacgcgtctctcccggtgtgtgactgcagacgaatgtgtagt 60
|||||

0v 61 tggcttgggtgtggcttattctgcagctggtgagatgtttatcaggcgcgagcgccatgaaggga 120
 Db 1333 ATGGTCTCTTCCAGATCTTCCGCTCTCTCCCGATATATGATCTCGACGATACCAATGATGATGACCA 2200

Db 1533 TGGCTGGGTGTCGCTTATTCTGCGGTGTTGGATGTTATCAGGGCAGCGGCGCATGAGGA 1474

Qy 121 gttacatagaaccgaaagccaagggcgccctgtagtctctcagagaagtgtgataactac 180

[illegible]

Db 1413 AACCTACACAGACGATCTAAGCGCGAGACCGAGACGACGATCTGTTTGTACGCC 1354

QY 241 cgcacctggttcgcttcaggaatatgactacgtccgcgcgtccattgcatgaact 300

DB 1353 C6CACCTGGTTTGTCTTCAGGAAATTAAGACATCGCCGGGAGCCTA 1254

Db 1293 ACGACCAACGATCTCGGTGTCTGCGGCACTCCGTACAGTAG 1249

RESULT 2

ADRCOMPEN/c			
LOCUS	ADRCOMPEN	35935 bp	DNA
DEFINITION	Mastadenovirus b5 gene, complete genome		
			VRL
			08-APR-1996

ACCESSION M73260 M29978
VERSION M73260.1 GI:209842

SOURCE Human adenovirus 5 DNA.

ORGANISM human adenovirus type 5
viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus
1 (beta 1 to 2005)

AUTHORS	TITLE
Chroboczek, J., Bieber, F. and Jacrot, B.	The sequence of the genome of adenovirus type 5 and its comparison

JOURNAL
OF
VIROLOGY
186 (1), 280-285 (1992)
WITH THE GENOME OF ADENOVIRUS TYPE 2
MEDLINE-
92087470

Location/Qualifiers

source	CDS
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

```

source
1..35935
/organism="Human adenovirus type 5"
/db_xref="taxon:28285"
<11565..12297
/note="52/55 K protein gene"
/codon_start=2
/protein_id="AA06406.1"
/db_xref="GI:209843"
/translation="SFNNHRTVAREEVAIGLMLMDFVSALQENPNSKPLAAOFL
IVOSRQNEAEADLLNITVEPEGKMLDNLNIIOSIVOCRSIADIKVAIVSMLS
LGKRYARIVTTPVDIKKEVIGFMDRLAKVLTISDDLGVRNRIKAVSRR
RELSDRELHMLORALAGTSGDREAEASYDAGADLMWPSRRALAEAGGPGLAAP
AAGAGVGEPEYDEDEPEPDCEY"
CDS
12318..14075

```

```

/codon_start=1
/product="Protein I11a"
/protein_id="AA96407.1"
/db_xref="gi:289094"

translation="MOQATDPVPAALQSPGSLNSTDMDQV64RMSLTANPAPA
PROQQAIRLSAIIELAVPAPARNPHTHEVILAIVALENRIIPDEGLVYDALLOQ
P/translation="

GDRALVSVLRKLVETEPQSEVYVQSGPDFTQSRGLOTVNLISQA FNKLQAGMISRAPTP
ITNFSALRLKLVETEPQSEVYVQSGPDFTQSRGLOTVNLISQA FNKLQAGMISRAPTP
GDRALVSVLRKLVETEPQSEVYVQSGPDFTQSRGLOTVNLISQA FNKLQAGMISRAPTP
GDRALVSVLRKLVETEPQSEVYVQSGPDFTQSRGLOTVNLISQA FNKLQAGMISRAPTP
EITSVSALRSGEDTSGLEATNLNLYLITNROKIPSLHSNSEEERILRYVOQSVLNLK
RDGVSPVSLALMTATANMEPGYASNPRIINLMYLIHAAAVNEEITNLNLINHWME
PSPGTEGFEVEEDNGDILMDIDDSVSPQPORLLEQOGEAEALRKESFRRSPSS
LSDGAAPRSDASSPSPLIGSLTSTPTTPRLGEEYVNNLSLLOPQRNLPAPAP
PNNGESLVYDCKSPKRYTAAOEHRDPGCRPPTRRQHRDRQGLVWEDDDSDADSVLLI
LGAGSPPFAHLERFRLGRNF"
16286..16474

```

```

/codon_start=1
/product="protein pVII"
/protein_id="AAA96408.1"
/db_xref="GI:209845"
/translation="MRAARLAAGIVTPPPSRRRRAAAAAAAAAAISAMTOGRGVNTWT
RDSVSGLRVPTRPDPNR"
16544..17650
/codon_start=1
/product="protein V"
/protein_id="AAA96409.1"
/db_xref="GI:209846"
/translation="MSKRIRKEMLOVTAPEIYSGPPKKEODYRPRKLRYAKKKKKD
DDEDDDEVLLIATAPRRVCQKGRVRYRATPTVTYTPGSESTPRYKRVYDEYD
DELDQDAERIGERFGRICRRKNDALIDEDENPPSLKPTVLOVLPALASERRR
LKRESDQLRIVQUMKPKRLEDYDEKATYVPGEGEYRNPVPIQVQAPGLGNOYD
QIPTSSISIAIATBGEHLOISPVASADAAVQVVAASASTSEVQDTPMPEFVS
PPRPSRSTISAGASALPEIATPASPAPGAGTIGTIRNRNATIRKRTITGTRRRR
RKPVLAIPISVRKRAVEGGRTVLTPFARIHPSIV
17774..17920

```

```

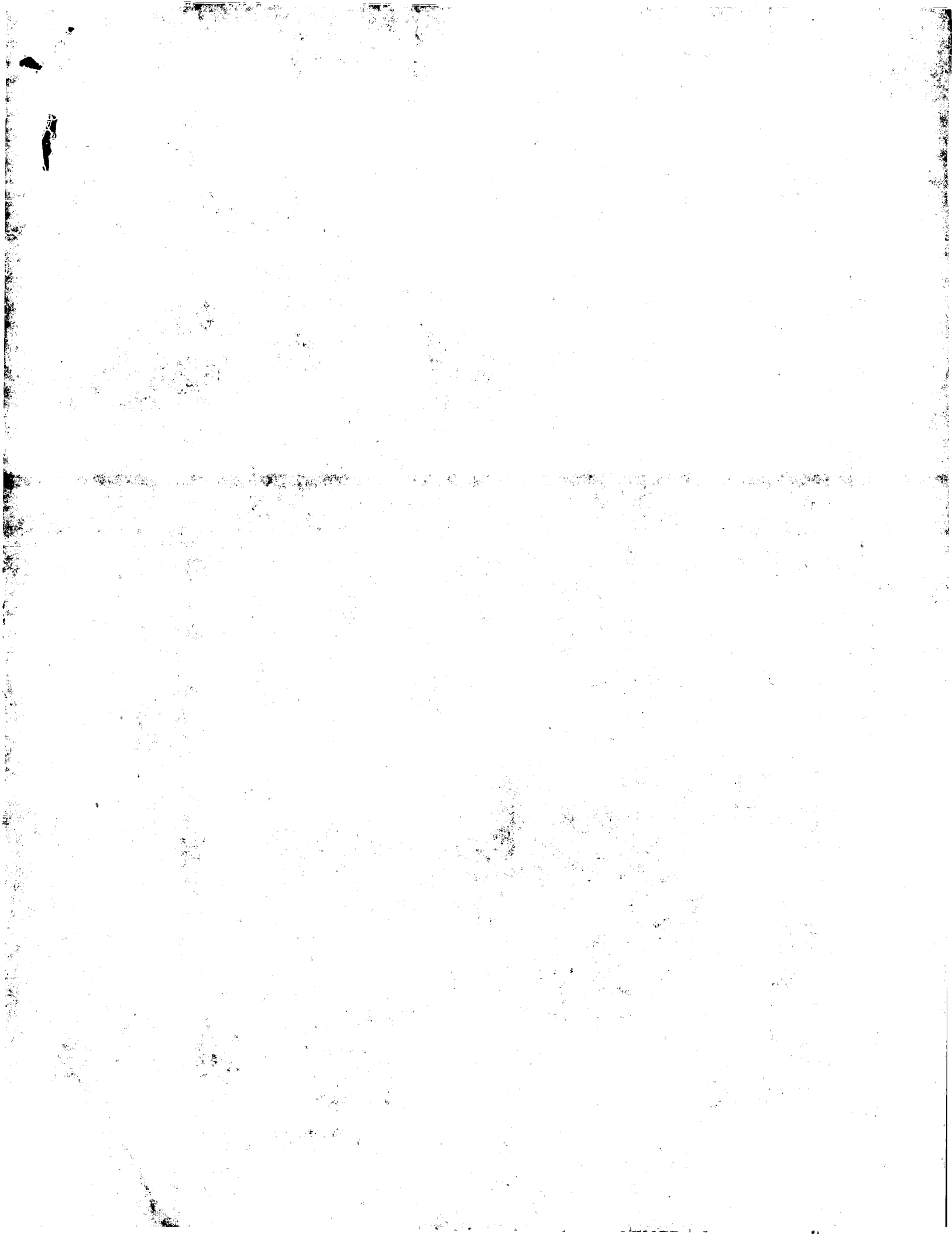
/codon_start=1
/product="protein 5 precursor"
/protein_id="AA96410.1"
/db_xref="GI:209847"
/translation="NRRAHHRRRRASHRMKRGILPLILPLILAAIGAVPGIASVALT
AQRH"
18003. .18755
/codon_start=1
/product="protein pVI"
/protein_id="AA96411.1"
/db_xref="GI:209848"
/translation="MEDINPASIAPRHGSRPFKMGQDICTSMGSGAPSKSGSINSNG
KNGSITPKNNGSKANNSSTGOMLRDLKQNTQOAKVDIASISGVLAQAVQNS
INSLRPLRPVEEPPPAVEVSPGEGEKRRPRDRREITVLOTIDEPSSYEALKQGL
TTPRIEMATGVGQHTPVTLIDLPADTQOAKVGLPTAVVTRPSRSALSLRAASG
RSLRPVNASGWSQSTLSIVGLQVSLKRRRCF"
<25819. .26484

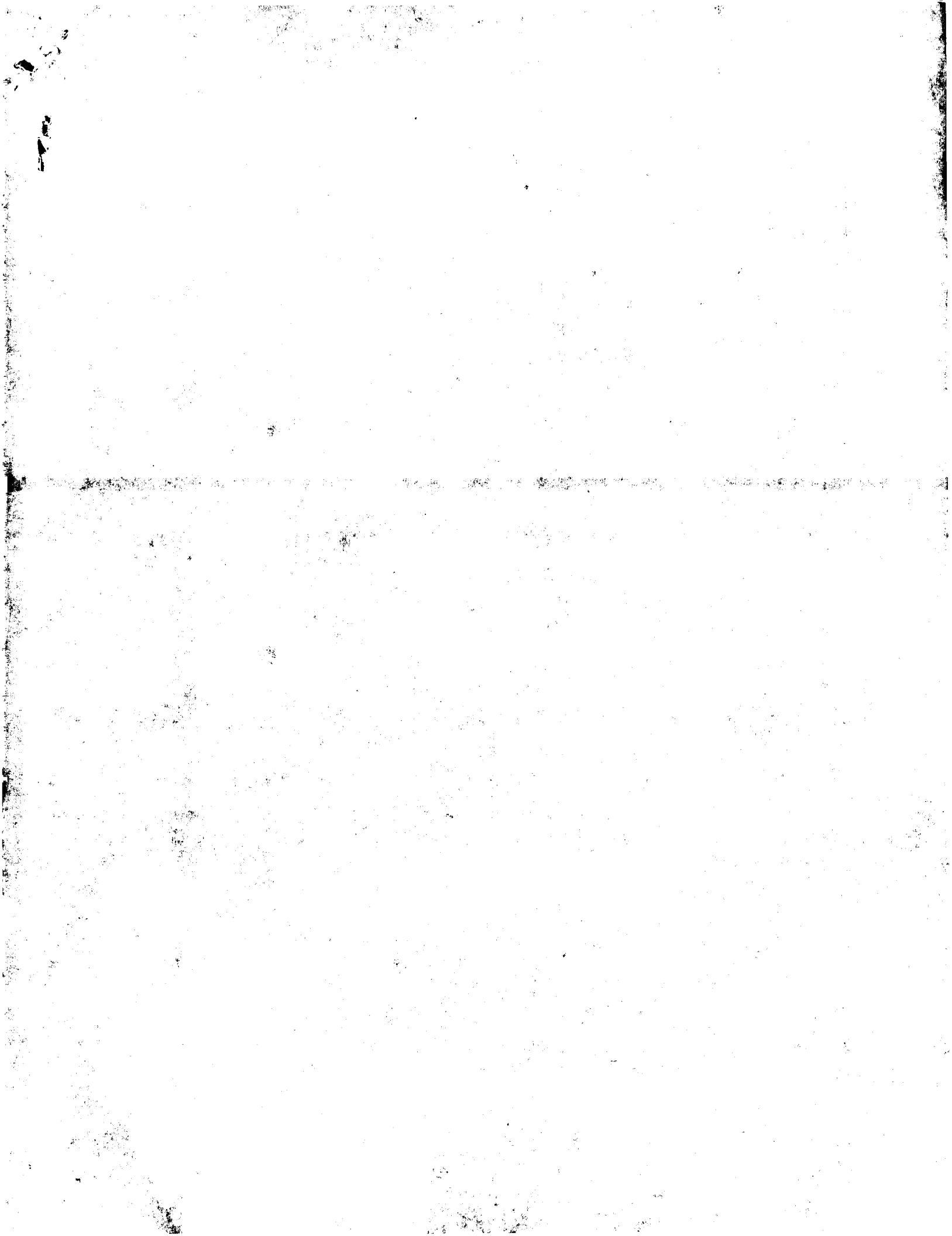
```

```

/note="100 k protein"
/codon_start=1
/protein_id="AA96412.1"
/db_xref="GI:209849"
/translation="ESQIETGFGSPDDEKSAFLKLTPTGLMTSAYLARKVDDP
HAHRIPEEDSSRPNAELTACVITGHTIGLQALINKAROEELTKRGSRVITPVED
ELNLTIPPPQYKQDPRLASQDGTOKELAAAAAATHGRGILIGSGRGFGGSGGG

```





```

source
1. .35935
/organism="human adenovirus type 5"
/db_xref="taxon:28285"
<11565 12297
/note="52/35 k protein gene"
/codon_start=2
/protein_id="AA96406.1"
/db_xref="gi:209843"
/translation="SFNNHVRFLVAREEVALIGLHMLDPFSALEQNPNSKFLMQLF
IYVSHSDNAPFADLNLVPEEGRWLDLNLITQSYVQERSLADKPAALNISM
LGFYFVRKIVPYVDIKKYEKGFTHRAALVTLSDGLVTRNERLHKVSVSR
RELSDELMHSQRLALACTGSGDREKAEISTFDAGADLRWPSRRALAAAGNGCLVAA
AARGNNGVEEIDEDDETEPEDGEY"
12318 .14075
CDS
/product="protein IIIa"
/protein_id="AA96407.1"
/db_xref="gi:289094"
/translation="MODADPAPVRAALQSGSLSTDDNQVDRITMSLTANPDA
FPOOQARLSSILEAVPARAPDTHKTKATLYNALAENRALRPDEGLVTDALLQR
AAYNSGNCOTNIDRLVGDVREAVAQERAAQOQSNLYNALAFSLSTPAPVPGQSD
YNTFSALRIKTEPQSEVYQSGPDFEFTQSNQGLQTNLSQAFNLQSGNIGRAVPT
GNATFVSLTPNSRLLLLIAPFTDQSGVSRDVTYGHLLTLVREAIQCAVHDHTQ

```

PPGFVYTGFEVPEBCNDGFLNDDIDDSVSPQOTLLELQQRQDAALRKESFSPSS
 ISLDIAALPVRDASPPSLIGSLTSTRTTBRLLGEEEYLNNSLLQQRKNTLPPAF
 PANGLESYDMSKMTYLAQEHNDVPGPPPTRRORHQRQGLWEEDSDSDSSVLD
 LGSGNPAAHLRPLGLGMF

CDS	16286.16474	/codon start=1
1	16286.16474	1

```

/product="protein pVII"
/protein_id="AA96408.1"

```

```
/db_xref="GI:209845"
/translation="MAARRLAAGIVTPPSRRRAAAAAAAAAISAMTQGRGNYTW
RDSVGLRVPVYTRPPRN"
```

```
CDS      16544..17650
/codon_start=1
```

```
/product="protein V"  
/protein_id="AA96409.1"  
/db_xref="taxon:30083"
```

```
/ou_xref="GI:209846"  
/translation="MSKRRIKEMLVIAVEIYGPCKEEDQYPRKLRYKKKKDD  
DDELIDDEFELIHATAADPDVPCDDPVRVVDGSSSTLHLLS
```

DEDLQANERLGEFAFGKRRKMDLALPDEGCPNTBSLPTVLOOVLPALAPSEERGLKRESGDLAPTVOLAYPKRROEDVLEKMTVEPGLFEPMBVIBTPVROUADGCIQUMG

QIPTSSSIATATEGMEJOTSPVASADVAQVAALAAKSTSEVQTDPMKRVSA
PRPRGSRKYGAASALLPEYALHPISAPTPGYRYTRPRRRATTBRTTTCGMBPB
PRPRGSRKYGAASALLPEYALHPISAPTPGYRYTRPRRRATTBRTTTCGMBPB

RQ0PVLAISVRVAREGGRTLVLPTRARHPSTIV
17774 . 17920

DS DS

```

/codon_start=1
/product="protein 5 precursor"
/protein_id="Protein5"

```

```
/protein_id=AA96410.1  
/db_xref="GI:209847"  
/translation="MPAHHHDDDSUPDHCCTTIT....."
```

/.....CLOO:
 AORH"
 18003. ,18755
 PDS

```
/codon_start=1
/product="protein pVr"
```

```

/protein_id="AA096411.1"
/db_xref="GI:209848"

```

Cribsiation-**MEDINASLAPRHSRPFMGKQMODIGTSNNSGAGPSWGLMGT
KNFGSVVKNYGSKAWSSTGMRLDKLEQNFOQRKYVDGLASISGVDDLANOAVQA
TNSKT.DDDDDRVREDDP.LHUGGGSSSSS**

INSGADLVNFEELEFPFAVELVSPGEGKGRKRPBDEETLWQIDEPSTYEALTKQGLP
TTRPLAPMATGYLGQHTPYVLDLPPPADTQOKPVLPGPTAAVVTRPSPASLRAASGP
RSLRPASGNGMSTINSTVIGVOSTREDDCE*

<25819..26484
/note="100 K protein"

```

/codon_start-1
/protein_id="AAA96412.1"

```

```

/db_xref="GI:209849"
/translation="ESQIGTFELQSPDEKSAAPGLKLTPLGLMTSAVLRKFVEDY
"

```

HAHLKE IEDOSNPNAELTACVITGHIIGLOAIINKAREFLRKRGVYLDPQSG
EELNIPPPQPYQQQPALASDQDQKEAAAAAATHGRGCIIGSGSGGFGRGSGCH

BASE COUNT	8367 a	10073 c	9761 g	7734 t
ORIGIN				

DB 3632 ATGGTTCCTCCAGCTCTTCCCGTCCCTCCCTGCTGACTCGCAGACGAATGCTAGGT 34283

61. tgcctggtggtgtgcttattctcgcggtggtgtaatgtaacagagcgcgcgcatgaagga 120
 121. |||||
 122. dbb_34482 TGCGTCGGGGATGGCTATATCTCGGCTGGGGCATGTTATCAAGGCGAGGCGCATGAAGGA 34222

Ob 3422 GTTACTAGAACCCAGACCAAGGGGGCCCTGGATGCTTGAGAGATGTATTACTAC 3416

Db 34162 AACCTACTACACAGGAGCTTAAGCGCGACAGACGACAGATCTGTTTCACGCC 34107

D03687.1
DB 34102 CGCACCCGGTTTGGCTTCAGGAAATATGACTACGTCCGGCGTTCATTGGCATTGACATGACACT 3404

Oy 301 acgacacacacgactctcggtgtgtctctgcgcgaactccgtaacgtag 345
 |||||
 Db 34042 ACAGACCAACGATCTCGGTGTGTCTCTGCGGCACATCCGTCAGTAG 33998
 .

RESULT 3
ADRCG/c

LOCUS	35937 bp	DNA	VRL	14-MAR-1996
DEFINITION	Adenovirus type 2, complete genome.			
ACCESSION	J01917	J01918	J01919	J01920
	J01921	J01922	J01923	J01924
	J01925	J01926	J01927	J01928
	J01929	J01930	J01931	J01932
	J01933	J01934		

J01995 J01936 J01937 J01938 J01939 J01940 J01941 J01942 J01943
 J01944 J01945 J01946 J01947 J01948 J01949 J01950 J01951 J01952
 J01953 J01954 J01955 J01956 J01957 K00086 K00394 K00395 K02367
 M3004 V00007 V00008 V00009 V00010 V00011 V00012 V00013 V00014

VERSION
KEYWORDS
V000015 V00016 V00017 V00018 V00019 V00020 V00023 V00024
J01917.1 Gi:209811
DNA polymerase; DNA-binding protein; RNA polymerase III;
alternative splicing; coat protein; complete genome; genome-11n

SOURCE
ORGANISM Human adenovirus type 2.
Human adenovirus type 2.

REFERENCE
1 (sites)
Ohe, K. and Weissman, S. M.
virtues; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovir

TITLE Studies of low molecular weight RNA from cells infected with adenovirus 2. I. The sequences at the 3' end of VA-RNA I

JOURNAL J. Biol. Chem. 252 (24), 9032-9042 (1977)

EDITING 78046048

REFERENCE	4 (sites)
AUTHORS	Celma, M.L., Pan, J. and Welsman, S.M.
TITLE	Studies of low molecular weight RNA from cells infected with adenovirus 2. II. Heterogeneity at the 5' end of VA-RNA I

JOURNAL
OF BIOL.
MEDICINE
78046049
5 (Doses 10514 to 10680)
Pan, J., Celima, M. L. and Weisman, S. M.
REFERENCE

STUDIES OF LOW MOLECULAR WEIGHT RNA FROM CELLS INFECTED WITH
ADENOVIRUS 2. III. THE SEQUENCE OF THE PROMOTER FOR VA-RNA 1
J. Biol. Chem. 252 (24), 9047-9054 (1977)
7804650

REFERENCE 6 (bases 30812 to 30900)
 AUTHORS Zain, B.S. and Roberts, R.J.
 TITLE Characterization and sequence analysis of a recombination site
 +the hybrid virus Ad2+Ad3

JOURNAL
OF MOL. BIOL. 120 (1), 13-31 (1978)
MEDLINE
78153757
REFERENCE
7 (sites)
Blundell C and Detterson II

AUTHORS
 KANISJAY, R. and REEDERSON, P.
 TITLE
 Nucleotide sequence at the junction between the coding region
 the adenovirus 2 hexon messenger RNA and its leader sequence
 JOURNAL
 Proc. Natl. Acad. Sci. U.S.A. 75 (12), 5822-5826 (1978)
 REFERENCE

MEDLINE	8 (bases 5986 to 6236)
REFERENCE	Ziff, E.B. and Evans, R.M.
AUTHORS	Coincidence of the promoter and capped 5' terminus of RNA
TITLE	

JOURNAL
MEDLINE
79084199
Cell 15 (4), 1463-1475 (1978)
adenovirus 2 major late transcription unit
9 (bases 21607 to 21816)

AUTHORS AKUSISIV, G. and FELLERSOHN, U.
TITLE Sequence analysis of adenovirus DNA. I. Nucleotide sequence
carboxy-terminal end of the gene for adenovirus type 2 hexon
JOURNAL Virology 91 (2), 477-480 (1978)

MEDLINE REFERENCE AUTHORS TITLE	
79119384	
10 (bases 1 to 156; 35804 to 35937)	
Arrand, J. R. and Roberts, R. J.	
The nucleotide sequences at the terminl of adenovirus-2 DNA	

JOURNAL
MEDICINE
REFERENCE
AUTHORS

J. Mol. Biol. 128 (4), 511-594 (1979)
7917534
11 (sites)
Zain, S., Sambrook, J., Roberts, R. J., Keller, W., Fried, M. and

Dunn, A.R.
Nucleotide sequence analysis of the leader segments in a cloned copy of adenovirus 2 fiber mRNA
Cell 16 (4), 851-861 (1979). 9

MEDLINE	79211249
REFERENCE	12 (bases 6039 to 6079; 7101 to 7172; 9634 to 9723; 18802 to 18891)
AUTHORS	Akusjarvi, G. and Pettersson, U.
TITLE	Sequence analysis of adenovirus DNA: complete nucleotide sequence of the DNA of adenovirus type 12

of the spliced 5' noncoding region of adenovirus 2 hexon mRNA
RNA
JOURNAL
Cell 16 (4), 841-850 (1979)
MEDLINE
79211248

REFERENCE	AUTHORS	TITLE
13 (sites)	Shinagawa, M. and Padmanabhan, R.	Nucleotide sequence at the inverted terminal repetition of

1

TITLE

Subject's name: [redacted]
[redacted] A.R.
[redacted] filed, M. and

